

China's 'eugenics' law still disturbing despite relabelling

As hoped, international discussions in Beijing on controversial family planning legislation have been productive. But China's law still gives grounds for worry.

Just under ten years ago, the Chinese province of Gansu introduced a local law graphically described in its initial English translation as "prohibiting reproduction of dull-witted, idiots or blockheads". For much of the intervening period, the Chinese government in Beijing has been attempting to draw up and put into practice a national law which, it claims, is aimed at the same public health objective, namely reducing the incidence of physical and mental disability, in particular through prenatal genetic screening. In doing so, China has raised considerable international hostility for efforts that some critics have even compared to those of Nazi eugenicists.

Much of this hostility has come from the genetics research community in the West, perhaps more sensitive than their Chinese counterparts to the abuses of their discipline in the past (including the recent revelations of enforced sterilization campaigns in apparently liberal countries such as Canada and Sweden). Professional societies in three countries — Britain, Argentina and Holland — registered their protest by boycotting the five-yearly International Congress of Genetics, which took place in Beijing last week (see page 711). Some individuals have suggested going further, by cutting off all research links with Chinese geneticists until the law, which came into force in 1995, is changed.

Ambiguities

One can have some sympathy for the Chinese defence that their strategy has been misinterpreted. Partly this was due to the naivety of an initial decision to describe the legislation, then in draft form, as a 'eugenics' law, apparently in ignorance of the negative connotations the word now carries in the West. Other misinterpretations appear to have been the result of ambiguities in the English translation of the final text of what is now known more benignly as the Law on Maternal and Infant Care. For example, it remains unclear whether a doctor's 'advice' on sterilization or ending a pregnancy is mandatory in cases where a "serious genetic disease" has been identified.

Furthermore, many of the law's articles about the need to provide adequate pre- and postnatal medical care for expectant mothers and children make good sense. Also, cultural differences in health-care delivery systems need to be respected. The traditional Chinese approach of treating a sick person not as an isolated individual, but as part of a family and a community, may not be aligned with the dominant views of western health-care providers; this does not mean, however, that such an approach is inevitably less effective.

But does this let China off the hook? Far from it. Even if the word 'eugenics' has been dropped from the official description of the law — and even if there is no explicit racist or political intent — it still contains elements and implications that are disturbing. One of these is the continuing lack of a clear definition of what is meant by "genetic disease of a serious nature". This omission can only increase the possibility of abuse. One Chinese bioethicist recalls being told by a local official in a remote region that three generations of mentally retarded individuals were sufficient to identify the syndrome as genetic — ignoring the possibility that the condition might be the

result, not the cause, of the poverty in which the family was living.

Similarly, the law remains ambiguous, even in its revised form, on the extent to which couples will in practice be left free of pressure to accept or reject a recommended sterilization. Doubts about this are only strengthened by a recent international survey showing that genetic counsellors in China — in common with those in many other developing countries, including India and Egypt — strongly favour what is called 'directive counselling'. Paternalist traditions remain strong in Chinese medicine, and few have the inclination to challenge their doctor's judgement.

Geneticists implicated

Two other elements in the current situation give pause for thought. One concerns the role of geneticists who, far from being ignored in drawing up the law, in fact played a key role in the process. Suspicion has been voiced in some quarters that, in their desire to distance themselves from the ideas of the Russian geneticist Trofim Lysenko, whose dismissal of genes as 'bourgeois ideology' dominated Chinese science for two decades, the geneticists may have ended up making over-optimistic promises about the extent to which their discipline is in a position to make an immediate contribution to public health.

Finally, there is the continuing danger that the provisions in the law could, without sufficient safeguards, allow genetic arguments to be used as a smokescreen for sanctions against potentially disruptive ethnic groups, for example Tibetan nationalists. There is no evidence yet that this has taken place; indeed those who accuse the Chinese authorities of using family planning rules against ethnic minorities often forget that they are applied more strictly to the Han majority. But, given concerns elsewhere about respect for human rights in China, and continuing nervousness within the country itself about allowing open debate on controversial topics, nothing should be taken for granted in the current circumstances.

Having said all that, the picture is not entirely discouraging. There is evidence that the Chinese authorities responsible for the law have been listening carefully to the criticisms of outsiders. This is particularly true of comments relating to apparent discrepancies between provisions in its text and the principles enshrined in various international codes of behaviour — such as the United Nations Universal Declaration of Human Rights — which China has agreed to sign as part of its entry ticket as a full member of the international trade community. Certainly, as far as geneticists are concerned, prolonged sniping over the law has cast an unwelcome shadow over some efforts to establish international collaborations.

The next congress of the International Genetics Federation takes place in Melbourne, Australia, in 2003. By then, it will be clearer whether fears about the implementation of China's new law have been exaggerated or justified. Meanwhile, the international debates stimulated by these fears have served to highlight some of the contentious issues that are surfacing as the techniques of genomic analysis become ever more sophisticated. Provoking reflection on these has been a positive step. □



Salk Institute investigated after claims of inhumane research

[SAN DIEGO] Federal authorities are investigating animal research at the Salk Institute, after past inhumane treatment and faulty experiments were exposed in a civil court case brought by a former employee.

The animal research programme at the institute in La Jolla, California, was so rife with problems at one stage that the success of scientific projects was threatened, the San Diego court heard.

Teresa J. Sylvina, a veterinarian hired in 1990 to correct animal research deficiencies at the Salk, is suing the institute for wrongful termination, retaliation, defamation and sexual harassment.

Most of the difficulties — ranging from deadly disease in animals to failed surgical techniques — occurred in the early 1990s, when the institute was trying to bring its programme into compliance with a 1985 animal care law. But internal strife over methods has lingered between scientists and administrative staff at the Salk, the court was told.

The federal investigation, launched last week, is being conducted by the US Department of Agriculture, which licenses animal research facilities, together with the Office for Protection from Research Risks at the National Institutes of Health (NIH).

Salk Institute officials say they have corrected the problems and now operate a first-rate animal research programme. They say they welcome the probe, which they anticipate will exonerate the facility.

Sylvina was fired in 1996, and now directs animal research at the Tufts University School of Veterinary Medicine in Boston, Massachusetts. But she told the court she was blackballed for more than a year by Salk Institute officials, in particular Nobel laureate Francis Crick, who once ran the institute on an interim basis.

The court heard about intense acrimony between Sylvina and Crick, who engaged in a "screaming match" in 1995. Crick couldn't be reached for comment last week.

Salk Institute officials contend that Sylvina was a bad manager. "There is no basis for the allegations presented by Sylvina in her lawsuit," said Fred H. Gage, chairman of the institute's faculty.

But some court records and sworn testimony support Sylvina's claims, and indicate that the institute's faculty was divided on her performance. Robert Hyman, a senior scientist at the institute who chaired the Animal Care and Use Committee, testified that, before Sylvina came aboard, "there was so much disease in the [animal] facility that people couldn't do experiments".



Troubled waters: Salk Institute animals were too sick for use in experiments, one scientist testified.

A major NIH cancer grant to the institute was threatened by the animal research deficiencies, Sylvina told the court. The institute's animal committee was "a dead-letter" panel, she said, which did not properly review or monitor experiments. Sylvina uncovered instances of inhumane and faulty experiments. But court records show the committee was reluctant to discipline scientists.

A neuroscientist who caused a number of problems, including accidentally burning one cat severely and asphyxiating another, was permitted to avoid a serious sanction, which would have been reported to NIH officials. The neuroscientist voluntarily halted his experiments and later left the institute.

When Sylvina was fired, Hyman testified that he resigned from the animal committee, believing she had been sacked for doing a good job. "She started with a failing facility there, she led a revolution there, which is never very comfortable for those in the status quo," Hyman testified. He declined to be interviewed for this article.

Federal authorities are also to examine claims of questionable NIH billing practices. Sylvina testified that she discovered a staff member of the animal department double-billing for experiments, but that administrators at the institute failed to aggressively pursue her concerns. A Salk administrator testified that the institute had given a monetary advance to the staff member, whose wages had been cut when Sylvina halted the questionable billing practice.

Rex Dalton

India set to allow patents for products

[NEW DELHI] India is to join the Paris Convention of the World Intellectual Property Organization (WIPO), in a move which is expected to bring a major liberalization of its patent laws.

The decision, which will simplify both international patenting for Indian scientists and the exploitation of international patents in India, was welcomed by scientists but criticized by some environmentalists.

"This is going to give our scientists tremendous impetus for research," says Ragunath Mashelkar, head of the Council of Scientific and Industrial Research, India's largest government scientific agency.

"It will lead to a lot of savings in the patenting process," says Suresh Chandran, an expert on patents at the National Institute of Immunology in New Delhi.

Under India's 1970 Patents Act, inventors of new chemicals or drugs can only patent manufacturing processes, not products. As a member of the World Trade Organization, India is already obliged to provide product patents by 2005. Changes to the 1970 act were resisted by the Bharatiya Janata Party (BJP) when it was in opposition, and last week's decision by the BJP government to join the Paris convention is seen as a signal that a change in the law is imminent.

Mashelkar says membership will allow Indian inventors access to WIPONET, a \$20 million computerized database on patents. India also becomes eligible to join the Budapest Treaty, allowing its scientists to maintain patentable microorganisms in their own repository instead of having to deposit them with one of the 26 repositories abroad recognized by the treaty.

Foreign scientists may now find the Indian patent regime less rigid, according to Chandran, because applications from member countries have to be given priority. Under the convention, a patent may not be refused because domestic law does not permit it — an indication that India plans to amend its law against product patents.

Vandana Shiva, an environmental activist, says that India will now come under pressure from multinationals to introduce product patents for drugs. She adds that drug prices rose by 20 per cent in Pakistan after it signed the convention in 1994.

Joining the convention before implementing biodiversity laws will, Shiva warns, "facilitate biopiracy globally and create a perverted situation, in which Indian patent offices will recognize patents based on biopiracy of Indian indigenous knowledge systems".

K. S. Jayaraman

Japanese R&D spending defies recession

[TOKYO] Despite being hit by its worst economic crisis since the Second World War, Japan's overall industrial research spending is rising steadily, according to a survey released last week by the Science and Technology Agency (STA).

The survey, based on data from more than 1,900 companies, says the total spending on research and development (R&D) for the 1997 fiscal year, which ended in March 1998, increased by 8 per cent from 1996, marking the third consecutive year of growth in industrial R&D expenditure (see graph).

The increase was centred on medium-sized and large companies with annual R&D budgets of more than ¥3 billion (US\$21 million), and on the transport and electronics sectors, which increased spending by 22 per cent and 12 per cent, respectively. Most companies with a budget of less than ¥3 billion have scaled down their R&D expenditure with the weakening of the Japanese economy, however.

The survey also says that Japanese companies have started to give more priority to R&D as part of their corporate strategy. Almost 90 per cent of companies surveyed have reorganized their R&D strategy in the past three years, of which 70 per cent —

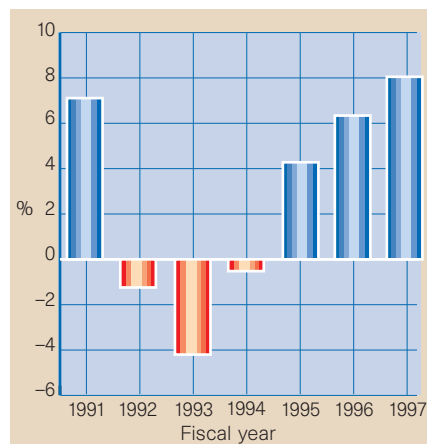
mostly in the transport, electronics and communications sectors — have made R&D their top priority.

This trend has been confirmed by a recent survey on corporate R&D by the daily business newspaper *Nikkei Sangyo Shimbun*, which indicates that more than 75 per cent of 442 companies surveyed claim to have increased their emphasis on R&D since 1995.

In both surveys, companies cite the development of innovative technologies and the creation of products that meet their market needs as the main reasons for raising the priority of R&D. Factors such as strengthening basic research and diversification of research areas were deemed less significant. This reflects the fact that companies' increased support for R&D is driven by their desire to strengthen application-orientated research.

Reiji Sano, managing director of Matsushita Electric Industrial Co., agrees that an increase in R&D expenditure has been accompanied by a shift towards applied research. "We are focusing our effort towards developing new technologies," he says.

STA warns in its report that the economy has deteriorated further since the survey was conducted early this year, and cautions that the actual increase in R&D spending for the



Rising sum: yearly increases in R&D spending.

full financial year may have been slightly less than that reported. But most companies say their R&D budgets remain relatively unscathed, despite the deepening recession.

According to a spokesman for the electrical manufacturer NEC, which increased its R&D budget by 9.4 per cent this year, the investment was planned on a 20-year basis and is unlikely to be affected by movements in the economy — "unless our profits really do plummet".

Asako Saegusa

Massive jobs shake-up as China reorganizes science academy

[BEIJING] Up to 20,000 researchers will be retired or transferred from their jobs within three years in the first phase of a dramatic reorganization of the Chinese Academy of Sciences. Only 10,000 research staff will be retained at the reformed institutes.

The president and five vice-presidents of the academy met in Huairou outside Beijing in the first week of August to thrash out details of the first phase of the planned reorganization of all the academy's 123 institutes and 68,000 researchers. The changes over 13 years will drastically reduce the number of full-time researchers and institutes and bring in new blood from outside (see *Nature* 394, 7; 1998).

It was decided at the meeting — but has yet to be officially announced — that in the first phase, lasting until 2000, 12 regional bases or centres will be created, or partially created, rather than the eight provisionally announced in June.

The additions include a centre for mathematical sciences in Beijing, one for astronomy in Beijing (linked to facilities in Xian, Nanjing, Shanghai and Yunnan) and small centres based on individual institutes in Nanjing, Dalian and Beijing. But plans for a south-western base of bioresources and biodiversity centred on Kunming have been postponed until a later stage.

The planned new regional centres	
Location	Field of research
1. Shanghai	Life sciences
2. Beijing	Mathematical science
3. Beijing linked with Shanghai, Xian, Nanjing, Yunnan	Astronomy
4. Beijing	Information science
5. Shanghai	High-technology R&D in electronics, materials science and other areas
6. Beijing	Physics/chemistry
7. Shenyang, Liaoning Province, north-east China	High-technology materials and advanced manufacturing
8. Beijing	Geoscience
9. Lanzhou, north-west China	Natural resources, environment, sustainable development
10. Dalian	Chemical physics
11. Beijing	Theoretical physics
12. Nanjing	Palaeontology

Institutes have been selected to participate in the first phase on the basis of having a proven academic track record, strong leadership, clear academic goals and a history of reform, academy officials say.

Under the initial reorganization, only about 10,000 of 30,000 researchers at the institutes involved in the first phase will retain full-time jobs in the academy. About 10,000 will be retired and another 10,000

will be transferred to other jobs, for example to companies affiliated with academy institutes. Selection of those to be retained will be a "tough job", admits a senior member of the academy.

For example, the 1,900 people employed at the eight institutes that will make up the Shanghai life-science research centre will be reduced to 800 researchers plus 100 administrators and technicians. The institutes currently include the Shanghai institutes of biochemistry, brain research, cell biology, entomology, physiology and plant physiology.

Directors of the new centres will be relieved of administrative duties that are unrelated to research, such as finding nursery schools for researchers' children or accommodation for staff. These tasks will be taken over by a small central administrative unit for a given group of institutes. The directors will assume roles similar to their counterparts in Europe or the United States, concentrating on research management and budgetary responsibility for the institute.

Advisory committees will select directors for the 12 new centres over the next few months. The directors will then choose group leaders, who will have the difficult task of deciding which of their researchers to retain.

David Swinbanks & Richard Nathan

Congress grabs eugenics common ground

[BEIJING] Scientists gathered at the International Congress of Genetics have warned against the use of genetic testing as a coercive tool of public policy. They call for all genetic counselling to be based on the principle of informed choice by individuals and their families.

A workshop on the science and ethics of eugenics agreed that governments should consult the scientific community on all genetics-related issues, and argued that the term 'eugenics' is currently used in so many ways "as to make it no longer suitable for use in scientific literature". Eight conclusions (see box) emerged from the workshop, which took place during the eighteenth International Congress of Genetics, held in Beijing under the auspices of the International Genetics Federation (IGF).

Although the topic of eugenics has been frequently discussed at the IGF's five-yearly congress, it attracted particular attention this year. This was because China adopted in 1994 what was initially called a 'eugenics law', which suggested that people suspected of having a 'serious genetic disease' might face compulsory sterilization if they wished to marry (see *Nature* 367, 3; 1994).

Chinese officials say that the main thrust of what has been subsequently renamed the Law on Maternal and Infant Health Care is a concern to provide a wide range of pre- and post-natal services. They say that concern outside China has been primarily the result of "ambiguities" in the wording of the section on genetic counselling.

Nevertheless such concern has led to considerable protest by groups ranging from US anti-abortionists to supporters of ethnic groups in Tibet. The Beijing meeting was boycotted by the professional genetics organizations of Britain, Holland and Argentina.

Last week's workshop was organized by the IGF as a dialogue with Chinese geneticists in a bid to stop the boycott from spreading further, and to placate those who wanted the meeting moved to another country as a protest. It was organized by Anthony Griffiths, professor of genetics at the University of British Columbia in Canada, who says that events in China have given the debate "a razor-sharp edge".

Two articles in the Chinese law have attracted particular attention. Article 10 says that if anyone intending to marry is diagnosed with "a genetic disease of a serious nature", the couple must agree to long-term contraception or sterilization. Under Article 16 if a physician "detects or suspects" such a disease in a married person, the couple should take "corresponding measures according to the physician's medical advice".

Qiu Ren-Zong, of the Chinese Academy of Social Sciences, argued at the workshop

'Informed choice must be the basis for testing'

Conclusions of the workshop on the science and ethics of eugenics, held during the 18th International Congress of Genetics in Beijing, 10–15 August 1998.

1. Countries share many ethical principles based on the will to do good and not harm. These can be applied in many different ways.
2. New genetic technology should be used to provide individuals with reliable information on which to base personal reproductive

choices, not as a tool of public policy or coercion.

3. Informed choice should be the basis for all genetic counselling and advice on reproductive decisions.
4. Genetic counselling should be for the benefit of the couple and their family; it has minimal effect on the incidence of deleterious alleles in the population.
5. The term 'eugenics' is used in so many different ways as to make it no longer suitable for use in scientific literature.

6. In formulating policy on genetic aspects of health, international and interdisciplinary communication should be carried out at all levels.
7. It is the responsibility of policy-makers concerned with genetic aspects of human health to seek sound scientific advice.
8. It is the responsibility of geneticists to educate physicians, decision-makers and the general public in genetics and its consequences for health.

that much of the controversy had been the result of the naïve use of the term 'eugenics' for the Chinese word *yousheng*, which he said can also mean 'healthy birth', with none of the negative connotations which the word has developed in the West.

Qiu emphasized that the law also has an economic function, in that China currently faced a significant bill in supporting its estimated 50 million disabled people. But he admitted that the "ambiguous" wording of articles 10 and 16 could be interpreted as giving doctors the power to require sterilization on the basis of a rudimentary diagnosis of particular symptoms being genetic in origin.

Several other speakers pointed out that the goals of the law could be given a more positive interpretation. Anne McLaren, for example, of Britain's Wellcome/CRC Institute at Cambridge, quoted one dictionary definition of eugenics as "pertaining to the protection of fine offspring", and said that many women in the West already take steps to achieve this. "We should either ban the term 'eugenic', or restore a definition in line with the Chinese meaning," she said.

Some speakers expressed particular concern at the way in which the law appears to give greater importance to the well-being of social groups than that of the individual. Qiu explained that this approach was not just a product of the socialist regime that has ruled China since 1949, but had been central to Chinese culture for many centuries.

Critics included Valery Soyfer, a prominent Russian agricultural geneticist and former dissident who now teaches at George Mason University in the USA. Soyfer argued that the Chinese position contradicted the United Nations Universal Declaration of Human Rights, which says that top priority should be given to the rights of individuals.

In response, Qiu argued that, even if the

meaning had been clouded by ambiguity in the language, one of the main themes of the controversial articles was the importance of what the West calls 'informed consent'.

"We must make clear that Chinese 'eugenics' is not the eugenics of Nazism, racism or genetic discrimination," he said. "Our only concern is to improve the health of the next generation."

Despite some sharp differences of opinion during the meeting, there was sufficient convergence of views — for example, on endorsing the concept of 'informed choice' as central to the use of genetic data in reproductive decisions — to lead to the decision to draw up a document expressing what Griffiths described as "a communal sentiment".

Presenting these to the final plenary session of the congress, he pointed out that the principles listed had implications that reached further than the issue of eugenics. "Geneticists have a broad responsibility to inform others about their science and its consequences for public health," Griffiths said. "If people do not fully understand what genetics is and is not capable of achieving, it is partly our fault."

The concerns of the meeting's critics appear to have been assuaged to some extent. Newton Morton, a British geneticist who chose to stay away from Beijing, describes the workshop's conclusions as "a splendid statement, and much more forthright than anything I had expected".

Morton, a professorial fellow in human genetics at the University of Southampton, says the conclusions were clearly "skillfully manoeuvred to win the support of the Chinese government to remove ambiguities in the law". But he adds that it remains to be seen whether the law is changed, or whether the workshop, and its conclusions, "rebound on Chinese geneticists".

David Dickson

Italy frees research council from 'baroni'

[MUNICH] The Italian council of ministers has approved sweeping reforms of the national research council, the CNR, that will reduce its influence over government science policy but may increase the efficiency of its research environment.

The ministers approved a decree introducing major changes to the law governing the CNR, which operates a network of 300 government research institutes in most disciplines of science. The changes are designed to curtail the pervasive influence over the CNR of an elite group of powerful university professors in Italy, known as *baroni*.

Scientists have cautiously welcomed the reforms, operational details of which have yet to be hammered out, but say that budget increases will be needed to allow the organization to fully exploit its new potential.

The power of the CNR president has been reduced and CNR's 14 discipline-based committees, which were heavily dominated by university professors, have been abolished. The committees had distributed research funds, appointed institute directors and served a pivotal role in the Italian government's formal scientific advisory system.

An executive board will be established consisting of the president of the CNR and six members, only two of which will be selected by the CNR. The board will make final decisions about institute directorships, which will have to be advertised openly.

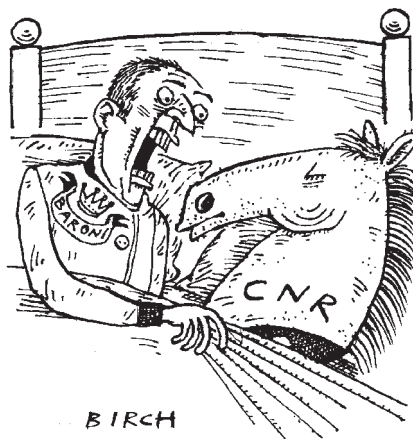
A scientific committee, comprising half CNR and half external scientists, will advise the executive board. It will authorize subcommittees to deal with research in specific disciplines, but the structure is designed to limit the influence of entrenched members of these subcommittees.

Administration will be decentralized, such that staff at the CNR headquarters in Rome could drop from more than 1,000 to under 300. The number of institutes and university-based CNR research centres will be reduced from 320 to under 100 by consolidating smaller centres into larger ones, but no loss of researchers is foreseen.

CNR will lose its grant agency role. Its small-grant fund, criticized for its 'raindrop' distribution (see *Nature* 367, 398; 1994), will disappear, and its strategic research fund will be transferred to a new ministry-based committee and opened up to scientists at universities and other research organizations.

Arturo Falaschi, a geneticist on leave from the CNR and currently director of the International Centre for Genetic Engineering and Biotechnology in Trieste, says the new CNR framework law "abolishes many of the faults of the past", particularly the way the CNR was used as a bargaining tool in academic politics by the powerful committees.

Falaschi says the autonomy given to insti-



tutes by the decentralization of administration will make the life of researchers simpler and more efficient. But true benefits will be reaped, he says, only if the new scientific committees are populated by good scientists committed to modern research and research management methods, including appropriate peer review.

Details of issues such as the recruitment of research staff and allocation of funds by peer review are to be drawn up by the CNR. How these details are defined is fundamental to the future of the CNR, says Falaschi.

Glauco Tocchini Valentini, director of the

CNR Institute for Cell Biology in Rome, says that if the details are not defined optimally, and if the government is not willing to raise the budget, which has stagnated for years through lack of political support, the CNR will be left very vulnerable. "We urgently need new blood, because only 14 per cent of our researchers are under 40," he says.

"If the government wants the CNR to build itself into the sort of internationally competitive organization it wants, it must provide adequate resources," Tocchini Valentini adds. Budget decisions for 1999 will be announced in the autumn.

The changes are intended to be instituted within 180 days, but Luigi Donato, director of the CNR Institute for Clinical Physiology in Pisa, thinks they are so extensive that two or three years could be required and that transition rules will therefore be needed.

The decree was made possible through the so-called Bassanini law, which allows existing laws to be modified without formal parliamentary approval to reduce the stalling powers of Italy's bureaucracy. It has taken more than a year to win government support for the reforms (see *Nature* 388, 609; 1997) because of strong opposition from the large cross-party lobby of politicians who are also university professors.

Alison Abbott

Boost to Canadian science infrastructure

[OTTAWA] Twenty-six Canadian universities have won infrastructure grants worth Can\$36 million (US\$24 million) under the first competition held by the Canada Foundation for Innovation (CFI). The CFI is an independent non-profit organization set up by the Canadian federal government last year to support research infrastructure.

More than 400 new faculty members will have their careers launched with buildings and equipment funded by the first CFI programme, called New Opportunities.

The CFI has Can\$800 million over five years to invest in infrastructure for research and development in Canadian universities, colleges, hospitals and other non-profit institutions (*Nature* 385, 759; 1997). It will typically provide 40 per cent of project costs, with the rest coming from local government, industry or the voluntary sector.

The response to its first competition, launched in December, has been overwhelming. Proposals worth almost Can\$3 billion were submitted, of which the CFI's share would have been Can\$1.2 billion. "All of us were quite flabbergasted," said David Strangway, the CFI president.

The contributions of other partners mean the CFI's Can\$36 million investment represents a total of Can\$90 million for

equipment and buildings. Strangway says the money will help young researchers recognized for tackling problems in priority areas for Canadians. These include understanding molecular processes in herbicide resistance, developing robotic systems for dynamic or dangerous situations, and building safer roads.

Robert Giroux, president of the Association of Universities and Colleges of Canada (AUCC), said the CFI was set up to end 10–15 years of serious underfunding of infrastructure that prevented Canada from capitalizing on emerging fields of research.

AUCC surveys show that better facilities, salaries and research funding abroad meant that many of the best researchers left the country, Giroux said.

CFI vice-president Carmen Charette said the criteria used to assess projects were research quality, infrastructure suitability, innovation capacity and potential benefit to Canada. She said New Opportunities would allow access to the best facilities.

Further competitions will be announced in autumn 1999 and the year 2000, and conferences in the next three to four years will consider future programmes. About half of the Can\$800 million will be available in the next two years.

David Spurgeon

NASA's Mars plans revitalized by Europe

[WASHINGTON] Increased European participation and new technical approaches are set to revitalize US plans for exploring Mars, which had been showing signs of becoming stalled by cost overruns and engineering difficulties, according to US officials. The new approach will transform NASA's plans for the exploration of Mars into a fully-fledged joint venture with overseas partners, say the officials.

NASA's decision in late spring to drop a long-distance rover from its planned 2001 Mars mission signalled that the agency's strategy to visit the planet roughly every two years and begin collecting samples in 2001 was starting to unravel. The rover had been designed to wander up to 100 kilometres from its landing site, gathering rock samples for return to Earth. But engineering development problems with the vehicle led to severe cost overruns that threatened to break NASA's \$200 million-a-year Mars budget.

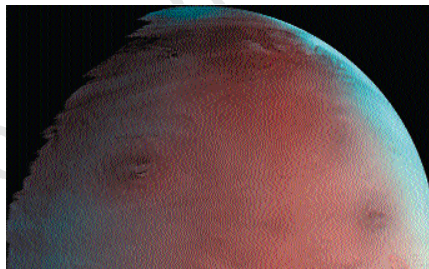
Now a far less capable rover is likely to fly on the mission, and no samples will be collected. Dropping the long-distance rover was a "considerable disappointment," says Michael Carr, a planetary scientist with the US Geological Survey in Flagstaff, Arizona.

This and other signs of trouble prompted NASA science managers to call for a fresh look at the long-term strategy for Mars exploration. A team of 30 engineers, planetary scientists, biologists and other specialists led by Charles Elachi of the Jet Propulsion Laboratory (JPL), California, has been meeting during the summer, and hopes by early September to produce a new draft

"architecture" for Mars exploration. Louis Friedman, director of the California-based Planetary Society, says the group has "a lot of clever people coming up with a lot of clever ways of doing things". As a result, he says, the Mars programme is "sparkling again".

Part of the reason is a growing European involvement with the programme. Elachi's team includes representatives from France, Italy and the European Space Agency (ESA). They bring engineering and scientific expertise, and the promise of financial resources beyond what the White House and US Congress have been willing to provide. Dan McCleese, chief scientist for Mars exploration at JPL, calls this a "significant new step".

"We had, up until this study, a NASA-only programme," McCleese says. "Now we've decided to build this [new] architecture assuming international cooperation." Key to the plan would be France's provision of an Ariane 5 launch for the sample return mission in 2005, as well as a Mars-orbiting spacecraft that would receive samples from the surface (see *Nature* 391, 213; 1998).



Sparkling again: international study group is bringing fresh impetus to plans for Mars missions.

The Italian Space Agency is proposing a communications satellite in Mars orbit to collect data from other spacecraft and relay them to Earth. ESA's proposed Mars Express mission, which hopes for a final go-ahead from the agency's Science Programme Committee in November and a launch in 2003, will carry an Italian-US radar system for detecting water beneath the surface of Mars.

France will also provide free 'piggyback' launches on Ariane, which can deliver additional small payloads (30 to 40 kilograms) to Mars at regular intervals. Elachi says these small missions — as many as four for every two-year launch opportunity — are likely to be an integral part of NASA's new plan as they could scout out sites for subsequent landers. By 2015 samples could be collected from up to six different locations on Mars.

One innovation that holds great promise is a recent proposal by JPL engineers to launch samples from the Martian surface to Mars orbit using solid-fuel rockets rather than heavier and more complicated liquid-fuel rockets. If approved, the technique could dramatically reduce costs or allow more scientific equipment to be sent to Mars.

The Elachi team will not deliver its final report until late September, but scientists who have been involved in the discussions are enthusiastic about the results so far. Steven Squyres of Cornell University, whose experiments on the 2001 mission had to be scaled down because of the loss of the more capable rover, says: "I'm feeling more optimistic about this programme than I have in years."

Tony Reichardt

Ethical protection for subjects 'could stifle psychiatric research'

[WASHINGTON] The leading US organization of medical schools is sharply criticizing a presidential commission's draft report that recommends new ethical protections for research subjects with mental health problems.

The Association of American Medical Colleges (AAMC) says the report's recommendations threaten to eliminate whole areas of psychiatric research. In a letter of 31 July to the National Bioethics Advisory Commission (NBAC), the association's president, Jordan Cohen, says the report paints an unfair picture of psychiatric researchers as disproportionately unethical. Cohen says that, by focusing on the mentally ill, the report risks stigmatizing people with mental disorders, when many other subjects can be said to have impaired capacity to consent.

The draft report, "Research involving subjects with mental disorders that may affect decisionmaking capacity," was

released on 1 July. It is the product of a year of work by the 17-member commission, which is considering the adequacy of current government research protections for the mentally ill.

In the late 1970s and early 1980s, two national commissions recommended that such protections be enhanced. But the Department of Health and Human Services shelved the recommendations after researchers complained that they would hamper research. There are echoes of those complaints in the current AAMC letter.

The association objects to a recommendation that "any apparent dissent" by a research subject must be honoured. If their surrogate representatives aren't allowed to overrule patients in some cases, it says, this "could make it impossible to complete research projects in Alzheimer's disease and other conditions".

Similarly, the association attacks a recommendation that research that does not

benefit subjects and that is of greater than minimal risk should be conducted only with the subject's informed consent plus the consent of a legally authorized surrogate. The AAMC says that requiring individual consent, coupled with the "vagueness" of the term "greater than minimal risk," could "effectively eliminate" research that does not benefit subjects.

But Stephen McConnell, vice-president for public policy at the Alzheimer's Association, says his group backs both NBAC recommendations. They "certainly will inhibit some research, but we will always defer to an individual," he says.

James Childress, the chairman of the NBAC subcommittee that drafted the report, declined to comment on the criticisms. He says the comments received will be taken seriously as the draft report is revised for presentation to the commission in mid-September. The text of the draft is at www.bioethics.gov.

Meredith Wadman

Johns Hopkins wins biotech patent case in federal court

[WASHINGTON] A US federal appeals court last week affirmed a lower court decision that a biotechnology company wilfully infringed on patents held by Johns Hopkins University of Baltimore, Maryland.

Johns Hopkins and two licensees had sued the Seattle company, CellPro, for patent infringement in CellPro's development and sale of a device that separates stem cells from blood and bone marrow. The invention underlying the device was discovered by Curt Civin, an oncologist at Johns Hopkins.

In July 1997, a federal district court upheld a jury verdict that the company had wilfully infringed university-held patents on the invention. Last week, the US Court of Appeals for the Federal Circuit in Washington agreed, opening the way for the university and its licensees to receive \$7.2 million in damages that the jury awarded. They are also asking for CellPro to pay \$8 million in legal fees. The lower court is expected to rule on these costs in the autumn.

The case drew broad attention in the biomedical research community last year when CellPro asked the government to exercise "march in" rights under a 1980 technology transfer law to force Johns Hopkins to license the invention to CellPro. Harold Varmus, the director of the National Institutes of Health, refused to do so.

UN biodiversity chief to step down from post

[LONDON] Calestous Juma, executive secretary of the United Nations biodiversity convention, is to step down when his current contract ends in October. The departure was a surprise to most of the convention's member countries.

The departure, which Juma attributed to personal reasons, will be a blow to the convention where Juma is held in high regard by both developed and developing countries. It will also affect talks on an international protocol on the safety of genetically modified organisms. The talks resumed in Montreal this week, and are at a precarious stage. Juma's successor has yet to be announced.

Scientist suspended in modified food scare

[LONDON] A leading research scientist was suspended, and is shortly to retire from his job, following his involvement in a controversial television documentary that claimed that genetically modified potatoes damaged the immune system of rats. The

documentary made headline news in Britain and reinforced concerns about the safety of genetically modified food.

The Rowett Research Institute, a government nutrition research laboratory in Aberdeen, suspended Arpad Pusztai, an authority on lectins, and withdrew its press release explaining the claim when a closer look at the data revealed that the experiment to which it referred had not yet been performed.

Human genome venture hires former NCI boss

[WASHINGTON] The company set up by Craig Venter, president of The Institute for Genomic Research (TIGR), and equipment supplier Perkin-Elmer to sequence the human genome more quickly and cheaply than the US government will be called the Celera Genomics Corporation.

TIGR has announced that Celera has hired Samuel Broder, a former director of the National Cancer Institute, as executive vice-president and chief medical officer.

The company has leased a 120,000 square-foot office building in Rockville, Maryland, two miles from TIGR, where 30 employees have been plotting strategy since early this month. The office building — now being partially converted to laboratories — will house 300 to 400 employees within two years, says Michael Knapp, a company spokesman.

Saxony goes to court to claim reactor costs

[MUNICH] The state of Saxony is to sue the German federal government over its decision, confirmed last month, not to help pay for the decommissioning of the former East Germany's only nuclear research reactor.

The reactor was built in 1957 at the Centre for Research in Rossendorf, Saxony. It was closed in 1991, shortly after the country's reunification, on the recommendation of Germany's science council, the Wissenschaftsrat. The research ministry of Saxony estimates the total costs of decommissioning to be DM400 million (US\$222 million) over the next decade.

The federal government has in the past contributed 80 to 90 per cent of the costs of decommissioning research reactors in West Germany. But it argues that it has no financial responsibilities for decommissioning nuclear facilities in east Germany.

A research ministry spokesman says Saxony will continue to finance the decommissioning as planned, but hopes that the courts will order the federal government to reimburse its costs.

Biotech pioneer to pay up over insider trading

[SAN FRANCISCO] Alejandro Zaffaroni, the biotechnology pioneer who founded four companies and headed the pharmaceutical corporation Syntex, has agreed with friends and family members to pay the US government \$1.85 million in returned profits and penalties for insider trading.

The charges stemmed from a flurry of profitable share deals that occurred just before Glaxo announced its acquisition of Zaffaroni's California-based company Affymax in January 1995.

Douglas Schwab, Zaffaroni's lawyer, said many of the purchases were made by family members who learned about the sale to Glaxo early because of a telephone call made by Zaffaroni, 75, to his 80-year-old sister explaining why he would have to miss her birthday party in Uruguay.

Three friends also bought large blocks of Affymax shares in the weeks before the deal was announced.

Zaffaroni was not accused of intentionally passing on inside information for his friends' gain or of trading illegally.

Indian astronomers reach deal on satellites

[NEW DELHI] India's radioastronomers have reached a time-sharing agreement with the US telecommunication company Iridium to protect their \$17 million Giant Metre-wave Radio Telescope near Pune from harmful radio interference from Iridium's satellites.

The agreement, reached after a year of negotiations, was announced on 13 August by Vijay Kapahi, director of the National Centre for Radio Astronomy. It means that Iridium is likely to obtain a government licence to launch a mobile phone service in India next month.

The deal follows similar agreements with US and European astronomers (see *Nature* 394, 607; 1998).

Fleming's penicillin mould goes under the hammer

[LONDON] A fragment believed to be grown



from the mould that Sir Alexander Fleming used to discover penicillin was sold at the London auction house Christie's last week for £8,050 (US\$13,000). The specimen (left) was sold in a circular dish inscribed by Fleming himself in 1954, a year before he died at the age of 74.

Serbia's universities come under attack

Sir—Recent political manoeuvring within the Serbian government in Belgrade has turned from bad to disastrous. The first target of the new coalition government has been Serbian universities and research institutes, followed by attacks on the independent media.

Top academic and scientific institutions have been chosen for various reasons. First, it is revenge for the universities' activities during the 1996–97 protests, which were instrumental in invalidating the mockery of an election. It is also a precautionary move against eventual new challenges to the autocratic regime. The second motive is the hatred which vice prime minister Dr (*sic*) Vojislav Sheshely and his followers feel towards academics and intellectuals in general, making the present attempts to destroy Serbian universities particularly dangerous. Sheshely is a member of the board of the Fund for Serbia's Development and of several university and institute boards.

Two main features in the University Act, passed recently in the National Assembly, threaten to have a devastating effect on the

educational system and scientific research.

First, the government has taken direct control of the entire structure and functioning of universities, appointing chancellors and deans, who in turn appoint practically all university staff. This has led to a purge of all "unsuitable employees", while incompetent people are being appointed to deans' and professors' chairs.

Second, all research institutes have been expelled from the universities (including eight from Belgrade University). The new minister for science and technology emphasizes that the institutes may only expect funds from the government if their products can find their way onto the supermarket shelves! The implications for fundamental research are clear. Scientific output, already badly hit by the sanctions, will drop rapidly.

As for the universities, the first practical step taken by the government has been to dismiss all the staff, and ask those who want to stay to sign "contracts" with new university authorities (a pledge of loyalty to the regime, in fact). Many employees have

refused to sign this humiliating paper and some have already been fired.

An immediate consequence of the new regime at the universities will be a drastic decrease in the amount and quality of teaching. In the past, some Serbian universities — in particular the largest, Belgrade — have had a good reputation abroad, and have provided gifted graduates to foreign research centres and colleges.

As a result of the current moves against academic freedom and top-rank intellectuals, most gifted young people will enrol in colleges abroad — if they can afford it. The most eminent professors will also leave the country for jobs abroad.

With the 'war' in Kosovo continuing, and the threat of military action against Serbia by the international community, prospects for the country appear gloomy indeed. As someone put it recently, with our current regime, we do not need NATO bombs at all.

Petar Grujic

*Institute of Physics,
PO Box 57, 11001, Belgrade, Yugoslavia*

Transgene risk is not too low to be tested

Sir—In the News story¹ "Organic farmer takes gene battle to court", the view is reported that "the probability of cross-pollination, and its effects on the environment and food safety, is too small to be studied effectively". But the probability of cross-pollination is a different issue from the effect of transgenes on the environment and food, which is not one of probability, but one that is amenable to experimental study. What is a matter of probability is whether cross-pollination can occur at all.

Studies of cross-pollination in maize in the United Kingdom are unlikely to be as extensive or informative as they are for some other well studied indigenous crops such as sugar beet. The probability of genes escaping in beet can be more accurately estimated^{2,3}, which may provide useful guidance for the maize problem.

Although quantification of rate of spread of transgenes through pollen and seed dispersal should be an essential part of risk assessment for genetically modified organisms, greater insight into the movement of genes can be gained by indirect methods. There are several documented cases of crossing having occurred between wild beet and sugar beet, evidence being derived from observations

of morphological genetic markers (see, for example, refs 4–6).

These results accord with a comprehensive assessment of other UK crops⁷ in which the average isolation distance for outcrossing species is calculated. Figures of 180 m and 3,200 m are quoted for maize and beet, respectively, but the important point is that actual distances of gene flow are almost certainly not only crop-specific but also variety-, site- and season-specific. As a result, the so-called isolation distance quoted in the News story of "about 200 m" for maize could allow as much as 10% of alien pollen to effect fertilization if work on natural populations of radish is a reliable indicator⁸.

Bearing these sources of inaccuracies in mind, can further scientific studies help to assess the risk of alien gene flow? It is clear that transgenic sugar beet is just as effective as non-transgenic beet at crossing with the wild type. "Weed beets", considered for many years to result from hybridization of sugar beet with wild beet, have been shown to be of such a hybrid origin using combinations of mitochondrial, chloroplast and nuclear DNA markers⁹. Microsatellite loci have been used to examine population structure among UK sea-beet populations¹⁰, pointing the way to their use for studying gene movement between the beet crop and wild forms. Other marker systems show that sugar beet and wild beet can be effectively discriminated¹¹, and DNA

microsatellite studies have demonstrated very high levels of polymorphism that could be used to discriminate between wild beets and sugar-beet varieties¹², with a high likelihood of being able to detect hybrids even of an introgressed nature.

The application of such markers is not confined to beets. Hybridization rates between wild *Brassica rapa* and cultivated *B. napus* have been reported in Scientific Correspondence¹³. Experiments such as these are essential to indicate how we can estimate risks of transgenes being released into the environment. They also call into question the view that the probability of cross-pollination is too small to be studied effectively.

Brian V. Ford-Lloyd

*School of Biological Sciences,
University of Birmingham,
Edgbaston, Birmingham B15 2TT, UK
e-mail: b.ford-lloyd@bham.ac.uk*

1. Masood, E. *Nature* **394**, 8 (1998).
2. Meier, F. C. & Artschwager, E. *Science* **88**, 507–508 (1938).
3. Artschwager, E. & Starrett, R. C. *J. Agric. Res.* **47**, 823–843 (1933).
4. Longden, P. *Pesticide Sci.* **7**, 422–425 (1976).
5. Evans, A. & Weir, J. *Kulturpflanze* **29**, 301–310 (1981).
6. Hornsey, K. G. & Arnold, M. H. *Ann. Appl. Biol.* **92**, 279–285 (1979).
7. Raybould, A. F. & Gray, A. J. *J. Appl. Ecol.* **30**, 199–219 (1993).
8. Ellstrand, N. C. & Marshall, D. L. *Am. Nat.* **126**, 606–616 (1985).
9. Boudry, P. *et al. Theor. Appl. Genet.* **87**, 471–478 (1993).
10. Raybould, A. F. *et al. Heredity* **80**, 127–132 (1998).
11. Shen, Y., Ford-Lloyd, B. V. & Newbury, H. J. *Heredity* **80**, 624–632 (1998).
12. Rae, S. J., Edwards, K. J., Burkholz, A. & Barnes, S. R. in *Plant and Animal Genome VI* Abstr. p. 72 (1998).
13. Scott, S. E. & Wilkinson, M. J. *Nature* **393**, 320 (1998).

Timing is everything in a game of two hemispheres

James W. C. White and Eric J. Steig

What causes ice ages, and other episodes of violent climate change? One clue comes from a new study of ice cores, which, unexpectedly, seems to show rapid warming events happening first in the Southern Hemisphere.

Leads and lags are at the heart of the palaeoclimate approach to understanding how the climate works. If event 1 leads event 2 in a palaeoclimatic record, for example, then event 1 could have caused event 2, but not vice versa. From this simple logic cause and effect is inferred, processes are proposed, models are tested, and our understanding of the climate system advances. But to establish leads and lags, dating of the records must be precise. For climate records extending back into glacial times, the state of the art has been to cross-date records to within about a thousand years. Yet many of the most dramatic climate-change events occur over much shorter timescales¹.

Blunier and co-workers, reporting on page 739 of this issue^{2,3}, have for the first time cross-dated isotope records in ice cores from Greenland and Antarctica to within a few centuries. As the ratio of ^{18}O to ^{16}O is a proxy for air temperature, this allowed them to establish more confidently the relative timing of large climate changes during the glacial period in the Southern and Northern hemispheres. It appears that the south leads the north by a little more than a millennium. A simple conclusion, but one that will ripple through the palaeoclimate and climate dynamics communities for some time.

The authors took advantage of the rapid and large swings in the methane concentration of the atmosphere during the last glacial period between 12,000 and 50,000 years ago. The fact that methane concentrations oscillate many times, each time nearly doubling in a few decades, is remarkable in itself, and the reasons for it are unclear. But Blunier and co-workers use these swings in methane, and the fact that the atmosphere is well mixed (so that a change in Greenland ice must coincide with a change in Antarctic ice), to put both ice cores on a common timescale.

First, they had to establish the difference in age between the ice and the gas it encloses, because the ice contains the stable isotopes that record climate, whereas the methane is trapped in air bubbles. As atmospheric gases

mix freely in the upper hundred metres or so of an ice sheet, the ice age at any depth is greater than the gas age. Where the accumulation of snow is very slow, such as at the Vostok base in Antarctica, the age difference is thousands of years and the uncertainty in dating is correspondingly large. But at sites with higher accumulation rates, the ice–gas age difference is much lower. By comparing methane at the GRIP site in central Greenland and the Byrd site in West Antarctica, Blunier and co-workers cross-dated the stable-isotope ratios in each core to within a few centuries. This is impressive when one considers that these events occurred 50,000 years ago at opposite ends of the Earth.

They find that during the last ice age, two rapid warming events of 15°C in Greenland occurred a millennium after smaller temperature rises in Antarctica. These are two of the ‘Dansgaard–Oeschger’, or D–O, warming events that are known to punctuate ice ages.

What does this mean? Have we found an

early warning system for climate change, and should we be watching the southern oceans for signs of warming? Although Antarctic temperatures have indeed warmed markedly over the past few decades, we should be cautious: climate changes during the ice ages, when the climate system is strongly affected by huge Northern Hemisphere ice sheets, may not occur in a similar way to those during interglacials. But the new work is important because it has sharpened our understanding of the global climate system, and moved us closer to the ultimate goal of predicting future climate changes. To see why, let us take a step back and put the new finding in context.

One of the most fundamental yet poorly understood facts about large-scale climate change in the past million years is that ice-age cooling and warming occur at the same time in both hemispheres⁴, to within a few thousand years. This seems odd, given that our best guess for the cause of ice ages is changes in incoming solar radiation in the Northern Hemisphere, and the concomitant changes in the Southern Hemisphere are opposite in sign⁵.

So what connects the climates of the two hemispheres? Is it the atmosphere, via greenhouse gases such as water vapour or carbon dioxide? Or the ocean, via deep currents of cold water formed near the poles? Blunier and co-workers argue that for the largest D–O events, where their measured time lag is more than 1,000 years, the ocean must be dominant — any atmospheric mechanism would operate much more quickly.

That brings us to another big question: do large climate changes, such as glacial cycles, originate in the tropics, which receive the bulk of the Sun’s energy, or at the poles,

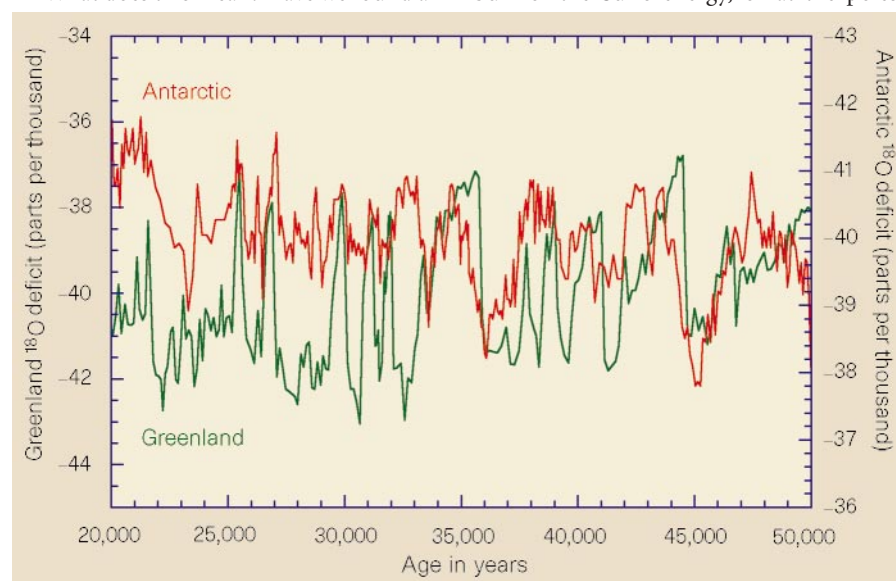


Figure 1 A comparison of the isotope records from the GRIP (Greenland) and Byrd (Antarctica) ice cores during the last ice age. The rapid oscillations, ‘Dansgaard–Oeschger’ events, are more pronounced in Greenland. The Byrd data have been inverted to show that a warming in Greenland may correspond to a cooling in Antarctica, an alternative to the hypothesis of Blunier and colleagues, who postulate that Antarctic warmings precede Greenland warmings by more than 1,000 years.

where the climate extremes are largest and there is a strong positive-feedback mechanism (as the region cools and ice spreads, more sunlight is reflected, lowering the temperature further)? The data collected by Blunier *et al.* indicate that these climate changes originate in the Southern Hemisphere. The authors offer an explanation of how the climate system can oscillate once it has been perturbed, but what starts the Southern Ocean warming? Here we are firmly in the unknown. In any case, this work re-focuses the attention of the palaeoclimate community on the Southern Hemisphere, after several years of intense scrutiny of the North Atlantic region, ignited and partly fuelled by the GISP2 and GRIP ice-core results from central Greenland.

Where do we go from here? First, how sure are we that there is a lead or lag present? Blunier and colleagues' dating seems solid, but the lag, about 1,500 years, is about a half-cycle for the D–O events, which tend to recur with an approximate, but poorly defined period of 3,000 years. This is, of course, the worst possible phase relationship, as it means that there may be neither lead nor lag, but simply an inverse relationship in climate change on this timescale between Greenland and Antarctica (Fig. 1). To get around this problem, Blunier *et al.* rely on peak matching, and assume that the two biggest D–O events in Greenland correspond to the two big isotope events in

Antarctica (see Fig. 1 on page 740).

Also, results from the an ice core at Taylor Dome, near the coast of eastern Antarctica, raise the possibility that the Vostok and Byrd cores are not entirely representative of Antarctic climate⁶. When the dating method of Blunier *et al.* is applied to Taylor Dome, the two largest rapid warming events turn out to be out of phase with those at Byrd and Vostok, but are apparently in phase with those in central Greenland. And results should soon be available from the tropical ice cores drilled and analysed by Lonnie Thompson at Ohio State University. How will these records fit with the polar results?

The conclusion may well be that the spatial distribution of climate change is highly textured, and that records from a much greater number of locations will be needed to complete the story. Timing may not be everything after all — these new results also highlight the importance of another cliché: location, location, location. □

James W. C. White and Eric J. Steig are at the Institute of Arctic and Alpine Research, Campus Box 450, University of Colorado, Boulder, Colorado 80309, USA.

e-mail: james.white@colorado.edu

1. Taylor, K. C. *et al.* *Science* **278**, 825–827 (1997).
2. Blunier, T. *et al.* *Nature* **394**, 739–743 (1998).
3. Blunier, T. *et al.* *Geophys. Res. Lett.* **24**, 2683–2686 (1997).
4. Broecker, W. S. & Denton, G. H. *Geochim. Cosmochim. Acta* **53**, 2465–2501 (1989).
5. Imbrie, J. *et al.* *Paleoceanography* **8**, 698–735 (1993).
6. Steig, E. J. *et al.* *Ann. Glaciol.* (in the press).

V(D)J recombination

Ragtime jumping

Ronald Plasterk

Transposons are generally considered the ultimate forms of selfish DNA — a single gene (or sometimes a set of two or three genes) that spreads simply because it ensures its own replication. Indeed, the ability of transposons to encode one or more proteins that selectively replicate the transposon is enough to explain their existence. But selfish elements may also end up doing something useful for their host, and it has often been speculated that transposon jumping may have generated gene arrangements that opened new avenues in evolution. Examples of this are rare, but a spectacular case has now been discovered. On page 744 of this issue, David Schatz and his colleagues¹ report that we owe the repertoire of our immune system to one transposon insertion, which occurred 450 million years ago in an ancestor of the jawed vertebrates. Vertebrates seem to have tamed this ancient transposon for generation of the immune repertoire, and the authors show that the RAG1 and RAG2 proteins (which mediate V(D)J joining) can still catalyse a full transposition reaction. A similar result has been independently obtained

by Martin Gellert and co-workers², and is reported in tomorrow's issue of *Cell*.

In developing lymphocytes, the crucial part of immunoglobulin genes — the exon that encodes the antigen-recognizing domain — is assembled from three seg-

ments, V, D and J, by deletion events referred to as V(D)J joining. Because several V, D and J segments can be joined, the joining process itself generates a combinatorial repertoire. Moreover, the joining step is 'sloppy', allowing for further extension of the repertoire³.

The V(D)J deletions are catalysed by RAG1 and RAG2 with the aid of several other factors involved in DNA bending and ligation. RAG1 and RAG2 were discovered by Schatz and Marjorie Oettinger when they were graduate students in David Baltimore's laboratory. They took a cell line that was not active in V(D)J joining, and introduced a stretch of DNA encoding an antibiotic-resistance gene. They had interrupted this gene with a segment of DNA that is typically deleted in V(D)J joining⁴, so the cell line was stably sensitive to the antibiotic. They then transfected in genomic DNA digested with a restriction enzyme, reasoning that if any of the restriction fragments encoded the gene essential for V(D)J joining, the transfectant could be selected — the recombinase would be able to delete the segment and generate an intact antibiotic-resistance gene. This unlikely schedule worked, and they identified the RAG1 gene⁵. The retrospective surprise became even bigger when it turned out that the restriction fragment actually contained two genes⁶, RAG1 and RAG2, both of which were required for V(D)J joining. Had those two genes not fitted within this one fragment, the selection would never have worked.

The V(D)J joining reaction was next reconstituted in a cell-free system by Gellert's group⁷ using purified components, and similarities between the chemical steps in V(D)J joining and transposition became clear. (This analysis profited from many years of study of recombination by Gellert and others at the National Institutes of Health in Bethesda.) DNA breakage and bonding were found to occur by nucleophilic attack of the DNA 3'-hydroxyl groups, in an energy-conserving fashion, without

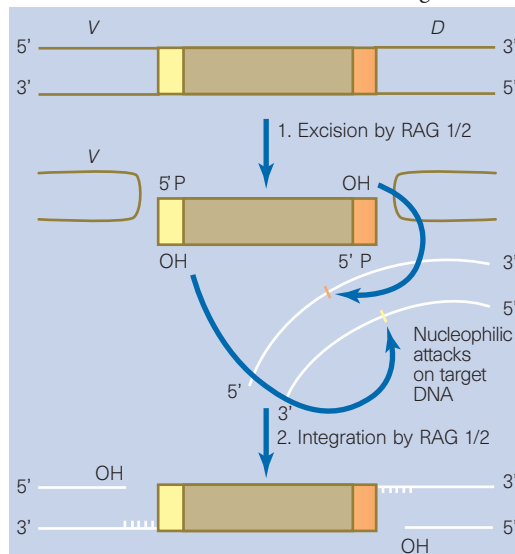


Figure 1 Transposition mediated by the RAG1/RAG2 transposase. The excision step is identical to the first step in V(D)J recombination; the second step is now reported by Agrawal *et al.*¹ and Hiom *et al.*². The combination of excision and integration is a complete transposition reaction, occurring by a similar mechanism to that used by other transposable elements¹³. After excision, the V and D segments have a hairpin at their ends, and can go on to be joined ('coding joint'). Owing to the staggered position of the bonds that are attacked in the target DNA, the integrated RAG transposon is initially flanked by five-nucleotide single-strand gaps.

covalent intermediates between DNA and transposase⁸. As well as the mechanistic similarities, the DNA sequences at the ends of the *Tc1* transposon of the nematode *Caenorhabditis elegans* were found⁹ to be very similar to the termini of the segment deleted in *V(D)J*, although the biological relevance of this was not clear.

The breakthrough now reported by Agrawal *et al.*¹ and Hiom *et al.*² started, as so often, by analysis of a thus-far overlooked product — a weak band on a gel which appeared as a side-product of the *in vitro* *V(D)J* reactions. Through a beautiful series of well-targeted experiments, they show that this product is the result of a transposition event. The products of RAG-mediated jumping show the hallmarks of real transposition — integration into more or less random positions, accompanied by duplication of a few (in this case five) base pairs of target DNA. This is due to the chemistry of transposition: the two transposon termini nucleophilically attack phosphodiester bonds in the two different strands of the target, five base pairs apart (Fig. 1).

The RAG proteins can mediate a complete transposition event — excision of what we might refer to as the RAG transposon, and reintegration into a new target. The second step does not seem to occur *in vivo* any more — all the better, because transposon insertions can be mutagenic, possibly carcinogenic. Instead, the excised transposon is immediately inactivated by joining of the two ends ('signal joint'). As pointed out by Hiom *et al.*, some classes of lymphoid tumour may still be attributed to partial transposition reactions by the RAG proteins. Knowing that RAG proteins can catalyse transposition, the sequence similarity with *Tc1* transposons⁹ seems to gain relevance. And, as pointed out by Spanopoulou *et al.*¹⁰, there are also similarities between the DNA-binding domains of RAG1 and *Tc1*-type transposases^{11,12}.

One of the wonders about the discovery of the RAG genes was that the two genes were both located within such a small segment of the genome. It is satisfying that Schatz can now provide an explanation for the luck that triggered his scientific career ten years ago: the two transposase genes are contained within such a small segment of DNA because they once had to fit in a small transposable element.

One might argue that such complex organisms as mammals (and other vertebrates) can only exist thanks to an immune defence system whose repertoire matches that of invading viruses and microorganisms (many of which, incidentally, use DNA rearrangements to increase their antigenic repertoire). If that is true, we may owe our existence to one transposition event that occurred 450 million years ago. The RAG transposon could have invaded one individ-

ual of the species that preceded the vertebrates and spread by transposition; then one insertion could have interrupted a membrane-spanning-receptor gene expressed in lymphocytes. The resulting inactivation of that gene would be reversible, through the action of the RAG transposase. Subsequent duplication events may then have built a large set of receptor-gene segments, which can now be recombined by RAG1/RAG2 in developing lymphocytes in several ways. As a result, the repertoire of somatic immunoglobulin genes can far exceed the number of immunoglobulin genes in the germ line. The reaction allows for an appropriate response to the almost infinite number of infectious agents that challenge us — in fact, we owe our life to the RAG transposase. □

Ronald Plasterk is at the Netherlands Cancer Institute, Division of Molecular Biology, Plesmanlaan 121, 1066cx Amsterdam, The Netherlands.

e-mail: rplas@nki.nl

1. Agrawal, A., Eastman, Q. M. & Schatz, D. G. *Nature* **394**, 744–751 (1998).
2. Hiom, K., Melek, M. & Gellert, M. *Cell* **94**, 463–470 (1998).
3. Tonegawa, S. *Nature* **302**, 575–581 (1983).
4. Schatz, D. G. & Baltimore, D. *Cell* **53**, 107–115 (1988).
5. Schatz, D. G. *et al.* *Cell* **59**, 1035–1048 (1989).
6. Oettinger, M. A. *et al.* *Science* **248**, 1517–1523 (1990).
7. Van Gent, D. C. *et al.* *Cell* **81**, 925–934 (1995).
8. Van Gent, D. C. *et al.* *Science* **271**, 1592–1594 (1996).
9. Dreyfus, D. H. *Mol. Immunol.* **29**, 807–810 (1992).
10. Spanopoulou, E. *et al.* *Cell* **87**, 263–276 (1996).
11. Vos, J. C. & Plasterk, R. H. A. *EMBO J.* **13**, 6125–6132 (1994).
12. Van Pouderooyen, G. *et al.* *EMBO J.* **16**, 6044–6054 (1997).
13. Plasterk, R. H. A. in *Mobile Genetic Elements* (ed. Sherratt, D. J.) 18–32 (Oxford Univ. Press, 1995).

Archaeology

Neanderthals emancipated

Paul G. Bahn

How human were the Neanderthals? Did they rival *Homo sapiens sapiens* in their art and technology, or merely copy us? Writing in *Current Anthropology*, Francesco d'Errico and colleagues¹ argue convincingly that the bone tools and ornaments manufactured by the Neanderthals at Arcy-sur-Cure in France, and elsewhere, represent an "independent development, not" — as has traditionally been believed — "the product of acculturation or imitation precipitated by cultural contact with modern humans immigrating into Western Europe".

One of the things that triggered this paper was a cover story in *Nature* two years ago² (Fig. 1, overleaf). J.-J. Hublin and colleagues showed that a juvenile temporal bone from Arcy, dating back to the Châtelperronian period (around 45,000 years ago), had come from a Neanderthal child. This, together with a Neanderthal skeleton from another French site (Saint-Césaire), found in association with a Châtelperronian stone industry, makes it reasonable to assume that Neanderthals created the Châtelperronian implements, previously thought to be the work of modern humans.

The fact that the Arcy Châtelperronian layers contained a wealth of ornaments and bone tools has led to all kinds of explanations. Neanderthals were thought to lack the intellectual ability to develop such modern features as personal ornaments. So, these artefacts were assumed to be imitations, made without understanding, or to have been gained by trade with, or scavenging from, modern neighbours. In other words, they were some ill-defined 'acculturation' from newly arrived superiors.

Attitudes to Neanderthals have always swung between extremes — take, for exam-

ple, the popular reconstructions in museums and books³, ranging from shaggy, sub-human brutes to creatures barely distinguishable from us. In the 1970s and 1980s, the view tended to be that not only were Neanderthals incapable of symbolic behaviour (such as the manufacture of ornaments or deliberate human burials), but that they were little better than animals. Only occasionally were voices raised against this bleak caricature⁴, sometimes underlining the existence of Neanderthal ornaments, including the specimens at the heart of the new paper^{1,5}.

The Grotte du Renne site at Arcy was excavated meticulously by French prehistorian André Leroi-Gourhan between 1949 and 1963. He encountered the Châtelperronian material, undisturbed and clearly separate from the overlying finds attributed to the Aurignacian period (about 40,000 years ago), which were presumably made by fully modern people. The Châtelperronian objects are clearly not the product of post-depositional disturbance because they are more numerous than the later specimens, and the layers in which they were found are a different colour. Nor were they acquired through trade or collection, as there are traces — the waste products of fabrication, for example — that they were made *in situ*. The artefacts do not closely resemble the Aurignacian specimens, so they cannot be the result of imitation. Moreover, the Châtelperronian bone and stone technologies were not learned from modern humans, because their production sequences (in, for example, the way blades were produced) are not identical to those of the Aurignacian, and they do not mark a distinct break with the preceding Neanderthal technological behaviour. ►



Figure 1 Pierced teeth and ivory rings from Arcy-sur-Cure². d'Errico and colleagues¹ argue that such ornaments were manufactured independently by Neanderthals.

d'Errico and colleagues¹ have now made a detailed reappraisal of the site and its contents. They conclude that "Arcy is consistent with the hypothesis of an original and independent cultural evolution of Western Europe's late Neanderthals". The Châtelperronian layers yielded 36 personal ornaments of different kinds — animal teeth, ivory beads and bones, showing perforations or grooves for suspension. There are also 142 worked bone objects of various sorts, most in excellent condition. Several are decorated with sequences of regularly spaced notches or incisions. The techniques used to perforate the roots of tooth ornaments were identical to those used at a different Châtelperronian site, Quinçay in France⁶, but very different from those known at Aurignacian sites. Similarly, no typical Aurignacian stone or bone tools have been found in the Châtelperronian layers, nor Châtelperronian ones in the Aurignacian. But artefacts from the Châtelperronian do show features that were already present in late 'Mousterian of Acheulian tradition' assemblages (a stone-tool industry in the Middle Palaeolithic, which is broadly equivalent to the Neanderthal period), well before the spread of the Aurignacian.

The Neanderthals of Arcy clearly had objects created for visual display on the body, and this implies the communication of some meaning. Indeed, d'Errico *et al.* claim that "Châtelperronian Neanderthals elaborated, used and transmitted autonomous codes, the reflexion of possible different social roles and the expression of a different cultural system". They conclude that Châtelperronian Neanderthals and Early Aurignacian modern humans were biologically different groups with important similarities in their cultural development. But we have no reason to assume that biological differences be-

tween them determined different intellectual capabilities, let alone that these would be detectable in the archaeological record.

The Neanderthals were cultured human beings, and we cannot assume that they were incapable of 'modern behaviour'. In fact, we have probably systematically underestimated their technological and symbolic sophistication. But opinions vary on the new study. Some feel that the authors have demonstrated their case beyond reasonable doubt; others cling to the traditional orthodoxy and question almost every fact, often contradicting their own earlier statements. One commentator, Margherita Mussi, makes a telling point — in Italy, Neanderthal sites with

'head-dominated' animal-bone collections have been interpreted as the result of scavenging from carnivore kills, whereas similar bone collections from Upper Palaeolithic sites (that is, Aurignacian or later) have been attributed to poor preservation and/or discard of 'non-diagnostic' bones by excavators and analysts. Such prejudice is also highlighted in a reanalysis of the large mammal bones from the Mousterian of Kobeh Cave, Iran⁷.

In their concluding remarks, d'Errico *et al.*¹ suggest that initial contacts between Neanderthals and modern humans triggered a rise in the use of symbols (that is, ornaments) on both sides — perhaps through the problems of personal, social and biological identity that arose out of such contact. In short, independent invention is a more satisfactory explanation than acculturation, with the two groups following parallel developmental tracks, and important social changes occurring in both. Their final sentence runs: "We believe that the time has come for the issue of Neanderthal extinction to be addressed with the tools of history and science rather than those of prejudice". Quite. □

Paul G. Bahn is at 428 Anlaby Road, Hull HU3 6QP, UK.

1. d'Errico, F., Zilhão, J., Julien, M., Baffier, D. & Pelegriin, J. *Curr. Anthropol.* **39**, S1–S44 (1998).
2. Hublin, J.-J., Spoor, F., Braun, M., Zonneveld, F. & Condemi, S. *Nature* **381**, 224–226 (1996).
3. Moser, S. *Antiquity* **66**, 831–844 (1992).
4. Hayden, B. J. *Hum. Evol.* **24**, 113–146 (1993).
5. Marshack, A. in *L'Homme de Néandertal* Vol. 5 La Pensée (ed. Otte, M.) 57–91 (ÉRAUL, Liège, 1988).
6. Granger, J.-M. & Lévêque, F. *C. R. Acad. Sci.* **325**, 537–543 (1997).
7. Marean, C. W. & Kim, S. Y. *Curr. Anthropol.* **39**, S79–S113 (1998).

Locomotion

Just skip it

Claire T. Farley

A girl and her dog are playing in the park. At first they walk slowly, side by side. As they speed up, the girl begins to run and the dog begins to trot. Finally, they break into a sprint and the dog begins to gallop. Whereas the girl uses the same running gait that she used at slower speeds, the dog has an extra gait, the gallop, that is unmatched by any normal bipedal gait. But in *Proceedings of the Royal Society*, Alberto Minetti¹ reports that skipping — the playful gait used by young children in the schoolyard — is biomechanically similar to quadrupedal galloping. In both skipping and galloping, pendulums and springs collaborate to conserve mechanical energy.

When people and animals walk and run, they use mechanical tricks to minimize how much work their muscles must do². During walking, they vault over their strut-like

limbs, conserving mechanical energy by using pendulum-like energy exchange. In the first half of the phase when a limb is in contact with the ground, the body's forward velocity slows as the body moves upward, and kinetic energy is converted into gravitational potential energy. During the second half of the contact phase for a limb, the opposite process occurs, as the body accelerates forward while moving downward. This recycling of mechanical energy reduces the work that the muscles must do by as much as 65%.

At faster speeds, people run and dogs trot. Both literally bounce along the ground, using springs rather than pendulums to reduce the work that their muscles must do. When a limb is in contact with the ground, its tendons and ligaments stretch and then recoil, storing and returning elastic energy³. It is difficult to work out exactly how much

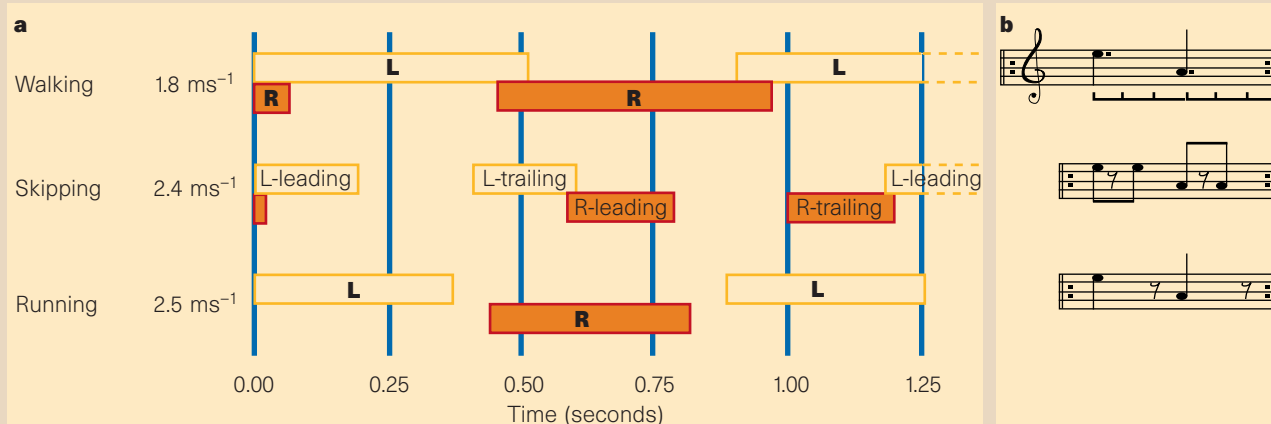


Figure 1 Patterns for walking, skipping and running in humans. Minetti¹ has shown that skipping in humans is also mechanically similar to galloping in four-legged animals. a, Contact patterns. L, left foot; R, right foot; leading, front foot in skipping; trailing, back foot in skipping. b, Rhythmic transcription. Walking and running show a two-beat pattern, whereas skipping shows a three-beat pattern. (Adapted from Minetti¹.)

elastic energy is stored in the body's many springs. In a single spring, such as a person's Achilles tendon, enough energy is stored and released to reduce the work that the muscles must do by 35% (ref. 3). Thus, muscles do only a fraction of the total work that is required for locomotion.

Gallop animals combine pendulums and springs to conserve mechanical energy². Now, Minetti¹ reports that humans also combine pendulums and springs in skip-

ping, a gait that young children adopt after they have learned to run⁴. Indeed, Minetti suggests that a galloping quadruped is like two skipping bipeds connected by a flexible trunk¹.

Skipping involves a flight phase, when neither foot is on the ground, followed by a ground contact phase, when the trailing foot and then the leading foot hit the ground in rapid succession¹ (Fig. 1). While the feet are on the ground, the fluctuations in gravita-

tional potential energy are approximately equal and opposite to the simultaneous fluctuations in kinetic energy. As a result, pendulum-like exchange reduces the work that the muscles must do by nearly 60%. High levels of pendulum-like energy exchange also occur when people use a bipedal 'galloping' gait⁵, a gait that is similar to skipping except that the same limb always remains in the lead.

Minetti uses a theoretical model to show that when people skip, elastic mechanisms may augment the pendulum-like energy-conserving mechanism. When the fluctuations in gravitational potential energy and kinetic energy are not simultaneous, tendons and ligaments can act as temporary energy-storage sites. For example, at the beginning of the ground contact phase of skipping, the body moves downward and then, a brief time later, accelerates forward. In this case, the gravitational potential energy that is lost early in the contact phase could be temporarily stored as elastic energy. Then, shortly after, the tendon could release the elastic energy, accelerating the body forward. In this way, springs allow exchange between gravitational potential energy and kinetic energy, even when their fluctuations are not simultaneous. Minetti's theoretical model indicates that this is what springs do in skipping.

If skipping involves such impressive mechanisms for conserving mechanical energy, why do people abandon skipping when they become adults? Four-legged animals routinely gallop over their whole life span, yet people skip only when they are young and playful. Galloping is the fastest gait for a quadruped, and the best in terms of avoiding injury and minimizing metabolic energy consumption at fast speeds. A four-legged animal can achieve faster speeds by galloping than by trotting because its back flexes and extends with each galloping stride, allowing exceptionally long strides. In

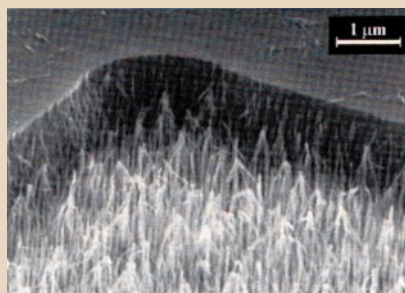
Semiconductors

Defective sculpture

Gallium nitride (GaN) is a semiconductor material that has great potential for use in optical devices such as green and blue light-emitting diodes. It is also riddled with structural defects. In the 10 August issue of *Applied Physics Letters* (73, 797–799; 1998), C. Youtsey, L. T. Romano and I. Adesida starkly expose these imperfections in a GaN film, by stripping away the intervening material. Characterization of the flaws may eventually aid the development of structurally pure films.

Resembling stalks of corn in a field, the thread-like defects, or dislocations, can be seen here, surrounded by untreated material. There can be no doubt that these structures coincide with dislocations — other images reveal that each thread stems from a 'root' defect in an underlying layer of untreated GaN film. But these delicate filamentary flaws are mere tens of nanometres thick. What kind of eraser could eliminate the material in between with such spatial selectivity?

The answer is a wet chemical etch process. A GaN film is first immersed in an alkaline solution in an electrochemical cell. Under illumination by ultraviolet light, positively charged holes ('missing'



electrons) are generated, which are involved in the oxidation and disintegration of the film. Yet somehow the defects are impervious to this electrochemical abrasion.

This resilience probably arises from electrical activity in the dislocations, which reduces the concentration of holes in the vicinity, suppressing the destructive oxidation reactions. The electrical activity may be caused by recombination of electrons and holes, or by charges at the dislocations that repel the holes.

Although it is a destructive technique, the erosion process could also be used to creative effect — it may be a way to make useful nanometre-scale structures in GaN films.

Karen Southwell

contrast, a person takes shorter strides and is slower during skipping than during running¹. Furthermore, quadrupedal galloping is a smooth gait⁶ that involves lower forces in muscles, tendons and bones than trotting at fast speeds^{7,8}. In contrast, skipping is a jolting gait that involves greater forces than walking or running¹. Finally, a quadruped consumes less metabolic energy to travel one kilometre at high speed by galloping than by walking or trotting⁹. In contrast, a person consumes more metabolic energy by skipping than by walking or running¹. The reason for the high metabolic cost of skipping, despite the tricks used to conserve mechanical energy, may be that the muscles must generate higher forces for skipping than for other gaits¹⁰.

In short, skipping is slow, jolting and tiring. This must be why most people abandon skipping when they become sensi-

ble adults. For children, however, the sheer joy of skipping seems to outweigh these rational considerations. □

Claire T. Farley is at the Locomotion Laboratory, Department of Integrative Biology, 3060 Valley Life Sciences Building, University of California, Berkeley, California 94720-3140, USA.
e-mail: cfarley@socrates.berkeley.edu

1. Minetti, A. E. *Proc. R. Soc. Lond. B* **265**, 1227–1235 (1998).
2. Cavagna, G. A., Heglund, N. C. & Taylor, C. R. *Am. J. Physiol.* **233**, R243–R261 (1977).
3. Alexander, R. M. *Elastic Mechanisms in Animal Movement* (Cambridge Univ. Press, 1988).
4. Clark, J. E. & Whitall, J. in *Development of Posture and Gait Across the Life Span* (eds Woolacott, M. H. & Shumway-Cook, A.) 128–151 (Univ. South Carolina Press, Columbia, 1989).
5. Caldwell, G. E. & Whitall, J. *J. Motor Behav.* **27**, 139–154 (1995).
6. McMahon, T. A. *J. Exp. Biol.* **115**, 263–282 (1985).
7. Farley, C. T. & Taylor, C. R. *Science* **253**, 306–308 (1991).
8. Biewener, A. A. & Taylor, C. R. *J. Exp. Biol.* **123**, 383–400 (1986).
9. Hoyt, D. F. & Taylor, C. R. *Nature* **292**, 239–240 (1981).
10. Kram, R. & Taylor, C. R. *Nature* **346**, 265–267 (1990).

Geology

Early uplift in Tibet?

William Ruddiman

On page 769 of this issue, Chung *et al.*¹ report evidence of potassium-rich volcanism about 35 million years ago in the northeastern Tibetan plateau, and infer that this region rose long before the southwestern part of the plateau. That has ramifications both for the tectonic history of the plateau and its suspected role in cooling the planet.

The Tibetan plateau covers 10° of latitude and 20° of longitude, more than a million square kilometres (Fig. 1). All of it lies above 5,000 m, high enough to bury the tallest peaks in the Alps and the Rockies. Its bedrock geology is still mostly unexplored, particularly in the north, and especially by techniques that constrain the timing of uplift. As a result, scientists often speak of the plateau as if describing a piston that has risen as a single unit.

In this picture, uplift was slow for the first 35 Myr (million years) after the initial collision between India and Asia, 55 Myr ago. But then, by 20 Myr ago, high topography had appeared. Its presence can be inferred from geochemical clues that show how long it takes minerals to move from hot regions many kilometres deep to exposure at the surface. This kind of evidence points to rapid 'unroofing' in the Himalayas and southern Tibet by 20 Myr ago, requiring rapid erosion, which in turn indicates that the surface was high². In addition, sediments eroded from the high Himalayas began to be deposited at slightly higher rates on the Indian Ocean floor around 20 Myr ago, and then at much higher rates ten million years later³. This also suggests that rapid uplift only took place in Tibet in the past 20 Myr.

But this evidence is biased towards the

fast-eroding high Himalayan mountains and the well-studied southwestern parts of Tibet. The piston history it provides omits the regional tectonic complexity that is typical of every continental region studied in detail. Continental collisions are inherently messy tectonic processes in time and space⁴, so we should expect uplift at different times in different parts of Tibet.

Chung *et al.*¹ move us towards this more detailed view. They find outcrops of potassium-rich volcanic rocks along several tectonic belts (faults and suture zones) crossing the northeastern part of the plateau,

mostly between 37 and 33 Myr old. Well-dated volcanic deposits of this kind in southwestern Tibet have already been used, together with signs of normal faulting, as evidence that the region has been under tension⁵ during the past 20 Myr. The pervasive extensional faulting is usually thought to mean that surface elevation is high — high enough to overcome the compressional stress applied by the collision of continents, allowing the plateau to sag under its own weight. Potassium-rich volcanism fits this picture because it is thought to result from melting in the upper mantle by hot asthenosphere (the more plastic part of the mantle) that has replaced cold lithosphere (the stiffer part of the mantle that usually underlies continental crust). For this replacement to occur, thick lithosphere had to be present in the first place, and then sink deeper into the mantle. Thick lithosphere points to high surface elevations.

By this chain of reasoning, Chung *et al.* infer that elevations in northeastern Tibet were high enough by 37 to 33 Myr ago to generate potassium-rich volcanism. If correct, this indicates much earlier uplift across some northeastern parts of the plateau than in the southwest. But it remains to be shown whether this early volcanism was merely localized near major faults, or was spread across a broad enough region of the plateau to typify much of its large-scale behaviour.

This new evidence also bears on the possible climatic role of the plateau, in the uplift-weathering hypothesis. It has been proposed that uplift of the plateau created and then strengthened the South Asian monsoon^{6,7}, and that monsoon rains worked in combination with exposure of fresh rock to



Figure 1 Tibet and the space shuttle's tail. The northeastern part of the Tibetan plateau (towards the lower left in this view) may have risen 15 million years before the better-explored south.

NASA/STS41G



100 YEARS AGO

When the year's work is over and all sense of responsibility has left us, who has not occasionally set his fancy free to dream about the unknown, perhaps the unknowable? And what should more frequently cross our dreams than what is so persistently before us in our serious moments of consciousness – the universal law of gravitation. We can leave our spectroscopes and magnets at home, but we cannot fly from the mysterious force which causes the rain-drops to fall from the clouds, and our children to tumble down the staircase. What is gravity? ... Lord Kelvin is quoted as having pointed out that two sources or two sinks of incompressible liquid will attract each other with the orthodox distance law. Let us dream, then, of a world in which atoms are sources through which an invisible fluid is pouring into three-dimensioned space. ... sinks would form another set of atoms, possibly equal to our own in all respects but one; they would mutually gravitate towards each other, but be repelled from the matter which we deal with on this earth. ... When the atom and the anti-atom unite, is it gravity only that is neutralised, or inertia also? May there not be, in fact, potential matter as well as potential energy? And if that is the case, can we imagine a vast expanse, without motion or mass, filled with this primordial mixture, which we cannot call a substance because it possesses none of the attributes which characterise matter, ready to be called into life by the creative spark? Was this the beginning of the world?

From *Nature* 18 August 1898.

50 YEARS AGO

The trouble in Palestine sets back the clock on the recent efforts of both Arab and Jewish gardeners to develop the horticultural attractions of the Holy Land, for the palm boulevards of Jaffa, and the flower-growing settlements at Mishmar-Hasharon, etc., had attracted much praise and attention. The danger, however, goes deeper, for modern Palestine was not the primitive wilderness of brigand and bedouin as depicted in most of the Western religious books. Several excellent gardens and plant collections were in the country, and their future is threatened by the bitterness of war.

From *Nature* 21 August 1948.

increase physical and chemical weathering, pulling more CO₂ out of the atmosphere and so cooling the global climate⁸.

The near-consensus of papers in a volume devoted to this topic⁹ makes the uplift-weathering hypothesis a leading explanation for global cooling during the past 20 Myr. Still, ignorance of uplift histories across much of Tibet has made the hypothesis difficult to evaluate in full. In particular, with little direct evidence for uplift before 20 Myr ago, it is hard to claim that Tibetan uplift caused, or was even involved in, the global cooling that began 55 Myr ago and led to Antarctic glaciation by 36 Myr ago¹⁰.

If large-scale uplift did occur in north-eastern Tibet as early as 37–33 Myr ago, chemical weathering of this high terrain could have contributed to global cooling then. Other evidence supports this idea. A striking increase in the global-ocean ⁸⁷Sr/⁸⁶Sr ratio began 40 to 35 Myr ago, and the extra ⁸⁷Sr probably came from the Tibet–Himalaya complex¹¹, from both accelerated weathering and the exposure of rocks rich in ⁸⁷Sr. The early uplift inferred for northeast Tibet

matches the initial upturn in this signal.

Now the question is whether further exploration of Tibet will find evidence of even earlier uplift, especially during the cooling between 55 and 40 Myr ago¹⁰. □

William Ruddiman is in the Department of Environmental Science, Clark Hall, University of Virginia, Charlottesville, Virginia 22903, USA.
e-mail: wfr5c@virginia.edu

1. Chung, S.-L. et al. *Nature* **394**, 769–773 (1998).
2. Harrison, T. M., Copeland, P., Kidd, W. S. F. & Yin, A. *Science* **255**, 1663–1670 (1992).
3. Rea, D. K. in *Synthesis of Results from Scientific Drilling in the Indian Ocean* (eds Duncan, R. A., Rea, D. K., Kidd, B., von Rad, U. & Weissel, J. K.) 387–402 (Am. Geophys. Un., Washington DC, 1992).
4. Molnar, P. *Am. Sci.* **77**, 350–360 (1989).
5. Turner, S. et al. *Nature* **364**, 50–54 (1993).
6. Ruddiman, W. F. & Kutzbach, J. E. *J. Geophys. Res.* **94**, 18409–18427 (1989).
7. Prell, W. L. & Kutzbach, J. E. *Nature* **360**, 647–652 (1992).
8. Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. *Geology* **16**, 649–653 (1988).
9. Ruddiman, W. F. (ed.) *Tectonic Uplift and Climate Change* (Plenum, New York, 1997).
10. Miller, K. G., Fairbanks, R. G. & Mountain, G. S. *Paleoceanography* **2**, 1–19 (1987).
11. Richter, F. R., Rowley, D. B. & DePaolo, D. J. *Earth Planet. Sci. Lett.* **109**, 11–23 (1992).

Making smooth moves

Terrence J. Sejnowski

Whether reaching, throwing, running or dancing, our natural tendency is to make smooth and precise movements. Out of the infinite number of ways that we could have made a particular movement, we generally pick the one that is the smoothest. The current thinking in the field of motor control is that the smooth, stereotyped trajectories made by our motor system are specially chosen to minimize jerkiness^{1,2} and to maximize efficiency³. Or could it be that smoothness is a by-product of a

more fundamental computational goal of the motor system, a goal that only makes us look graceful by accomplishing something else?

On page 780 of this issue⁴, Harris and Wolpert propose an alternative to the principle of maximum efficiency: the principle of maximum precision. On the face of it, making a precise movement does not seem to imply smoothness. Imagine that the goal is to touch an object as precisely as possible in a fixed amount of time. Getting to the spot as

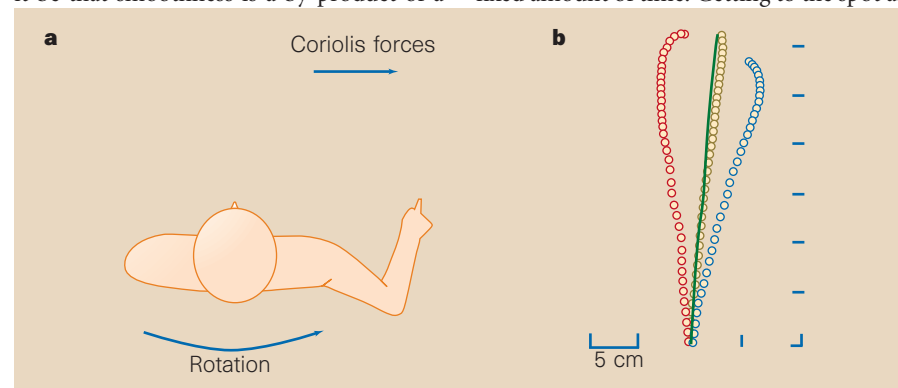


Figure 1 Hand trajectories for reaching before and after rotation of the body, showing that smooth movements are made even after adaptation in an altered environment. a, The subject was on a turntable and slowly rotated. b, View from above of average reaching movements, made in darkness to the position of a visual target that was extinguished just before the subject reached for it. The initial trajectories after the start of rotation (blue circles) were seriously affected by Coriolis forces, but after 40 arm movements (yellow circles) the accuracy and velocity profile of the trajectory was almost identical to that of the original movement (green line). After the rotation stopped, the initial trajectory (red circles) shows the after-effects of rotation. (Adapted from ref. 7.)

quickly as possible and making a careful landing might be a better way to ensure precision rather than making the arm movement as smoothly as possible. What this high-acceleration strategy does not take into account, however, is that the motor neurons that control the arm are noisy and cannot be counted on to get the arm to the same place for the same command. In particular, giving your muscles strong commands is exactly the wrong thing to do because the variability in the muscle output increases with the strength of the command. When they take activity-dependent motor neuron noise into account, Harris and Wolpert find that optimizing the precision of the endpoint of a movement produces smooth movements with exactly the properties of those that we tend to make.

The velocity profiles of arm movements are highly symmetrical around the midpoint of the movement. Simulations of a simple arm controlled by signals that have the same noise characteristics as our motor neurons have similar bell-shaped trajectories when the controller is optimized for maximum precision, over a wide range of parameters such as the inertia, viscosity and time constants. Robustness of the shape of the velocity curve to the details of the arm model is particularly important because the universality of this property⁵ originally inspired the minimum-jerk model of motor control. Not only does smoothness apply to arm movements under normal conditions, it also holds after adaptation in altered environments (Fig. 1). The maximum-precision model also accounts for other universal laws of arm movements, such as those that relate the duration of a movement to the maximum precision attained⁶, and how the speed of movement scales with the radius of curvature. It is also impressive that the model can account for a broad range of movements including saccadic eye movements, pointing movements and rhythmic arm movements.

This explanation is satisfying for three reasons. First, reducing uncertainty should clearly be a primary concern of any movement controller, whereas smoothness might reduce wear and tear but is more of a luxury. Second, the motor system is constantly calibrating itself to improve performance⁷ (Fig. 1), and it is much easier to compute the endpoint error than the degree of smoothness. Finally, grace is a reward for virtue, a bonus for being as accurate as you can possibly be. So we may move through the world smoothly neither by chance nor necessity, but rather because of noise in our motor neurons.

Noise is ubiquitous in the nervous system and is often ignored. The response of a neuron to a sensory stimulus or the output of a motor neuron during an action is highly variable from one experimental trial to the next, so responses are typically averaged over

dozens of trials to reduce the variability. However, we can see clearly and move accurately on a single trial, so it is of interest to look more closely at the limits and possible benefits of neural variability. The existence of invertebrate brains that work at a much higher level of precision and repeatability with many fewer neurons, and the exceptional precision in the timing of spikes in the peripheral auditory and electrosensory systems of vertebrates, suggest that noise is not an inevitable consequence of sloppy components.

To a first approximation, the variance in the firing rate of a neuron is proportional to its rate; in the cortex of the brain, the ratio of the variance to the rate is close to 1, making the spike trains of cortical neurons about as variable as radioactive decay. One possible benefit of having this degree of variability is to keep a neuron poised at its most sensitive region, near the threshold and ready to fire a spike whenever a suitable excitatory signal appears⁸. A neuron that fires with a high degree of variability can carry more information than one that fires at a constant rate, like a metronome⁹. But it is not yet clear how brains take advantage of the bandwidth in the spike timing. There may be other reasons

for neural variability that we do not yet fully appreciate. The observation that, because of noisy motor neurons, we may have to move more smoothly in the outside world, could have a counterpart for internal brain functions that also tend to run smoothly^{8,10}. Noise may not be a problem for neurons, but a solution. □

Terrence J. Sejnowski is with the Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA, and the University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA.
e-mail: terry@salk.edu

1. Hogan, N. *J. Neurosci.* **4**, 2745–2754 (1984).
2. Uno, Y., Kawato, M. & Suzuki, V. *Biol. Cybernet.* **61**, 89–101 (1989).
3. Nelson, W. L. *Biol. Cybernet.* **46**, 135–147 (1983).
4. Harris, C. M. & Wolpert, D. M. *Nature* **394**, 780–784 (1998).
5. Atkeson, C. G. & Hollerbach, J. M. *J. Neurosci.* **5**, 2318–2330 (1985).
6. Meyer, D. E., Smith, J. E. K. & Wright, C. E. *Psychol. Rev.* **89**, 449–482 (1982).
7. Dizio, P. & Lackner, J. R. *J. Neurophysiol.* **74**, 1787–1792 (1995).
8. Sejnowski, T. J. in *Explorations in the Microstructure of Cognition*, Vol. 2 Applications (eds McClelland, J. & Rumelhart, D.) 372–389 (MIT Press, Cambridge, MA, 1986).
9. Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, 1997).
10. Postma, E. O., van den Herik, H. J. & Hudson, P. T. *Int. J. Neural Systems* **7**, 537–542 (1996).

Pollination

Sunbird surprise for syndromes

Jeff Ollerton

Interactions between plants and their pollinators include some of the most striking and sophisticated of ecological affiliations¹. On page 731 of this issue, Anton Pauw² describes the relationship between the South African plant *Microlooma sagittatum*, a member of the milkweed family, and its pollinator, the lesser double-collared sunbird (*Nectarinia chalybea*; Fig. 1). The importance of this work lies not only in the description and experimental demonstration of a surprising interaction — it also highlights the danger of assuming that a pollinator is known from an analysis of a pollination 'syndrome'.

Pollination syndromes are suites of flower characteristics (morphology, colour, nectar and odour) that supposedly attract particular pollinators to specific flowers, and allow them to forage at the exclusion of 'illegitimate' visitors that would take the floral reward without executing pollination^{1–3}. This idea is superficially tidy, and it appeals to the classifying minds of many biologists. Indeed, some authors have used this concept to infer the pollinators of species even without field data and then to draw far-reaching conclusions about the historical ecology and evolution of such relationships⁴.

But there are problems associated with the syndrome approach. First of all, most

flowering plants are pollinated by a wide taxonomic range of pollinators⁵, and cannot be shoehorned into neat syndromes. Second, the jaw-cracking terminology of the syndrome concept is often imprecise and confusing. For instance, the luridly coloured, foul-smelling flowers associated with 'sapromyophily' (literally, flowers that mimic decaying organic material and are 'loved' by flies) are often beetle pollinated. An example is the now famous *Amorphophallus titanum*⁶, one of which flowered this June at Miami's Fairchild Tropical Garden. Finally, and surprisingly (given their wide acceptance), the predictive value of pollination syndromes has never really been tested — there has been little attempt to use these descriptions as hypotheses to verify the usefulness of floral characteristics in predicting what might pollinate a particular plant. The syndrome of 'ornithophily' (bird pollination), for example, is typified by tubular, red, scentless flowers. Although many ornithophilous flowers fit this description, many do not; hummingbirds, for instance, can visit a range of flowers, regardless of morphology or colour⁷.

Some workers have recently begun to take a healthily sceptical approach to the syndrome concept^{5,8}. However, there are no published examples in which communities



Figure 1 Probing for pollen — the lesser double-collared sunbird *Nectarinia chalybea*. Anton Pauw² has found that *N. chalybea* pollinates one member of the milkweed family, *Microloma sagittatum*, by transferring pollen between plants on the tip of its tongue as it probes for nectar. This association is surprising because, from its morphology, *M. sagittatum* was not predicted to be bird pollinated.

of plants have been classified according to their pollination syndromes and then surveyed for the actual pollinators. But I suspect that such an exercise would result in the correct prediction of only a small number of species from some of the more extremely specialized syndromes (such as ‘sphingophily’ — hawkmoth pollination). Most species would be difficult to classify. Tellingly, the main pollinators of many species vary between seasons and populations⁸, making it difficult to infer the adaptive regime that is experienced by plants in the face of temporally and spatially variable natural selection⁹.

Pauw’s finding — that *M. sagittatum* is bird pollinated — strengthens this argument because it goes against all predictions. Although some South African flora are bird pollinated, these plants usually dispense powdery pollen which dusts the plumage of the visiting birds. But *M. sagittatum* is a typical member of the family Asclepiadaceae and, as such, it does not possess free pollen. Instead, it has compact accretions of pollen that attach to the pollinator through mechanical clips (‘pollinaria’). These pollinaria are tiny (typically 0.25–1.25 mm in length¹⁰), and were previously thought to take part only in insect pollination¹¹. As Pauw now describes, however, the pollen masses attach to the tongue tips of the sunbirds as they probe the flowers for nectar, and can be transferred between flowers in this way.

The short floral tube, small entrance holes and, perhaps most significantly, presence of pollinaria previously excluded *M. sagittatum* from being recognized as a bird-pollinated plant² — these characteristics did not fit preconceived ideas of what such a plant should look like. Based on a comparative survey of the floral traits of *M. sagitta-*

tum, Pauw suggests that six of the nine species of *Microloma* ought to be bird pollinated. Further field observations and experiments are required to test this prediction because, as Pauw himself has shown, knowing the flower may not be enough to know the pollinator.

Previously unknown examples of plant–pollinator interactions are constantly being reported, and I would not even be surprised to read of a fish-pollinated aquatic plant. However, these relationships are more than just entertaining diversions — pollination is an ecological process that is vital to the preservation of diverse plant communities¹², so it may indirectly affect local primary production and ecosystem stability¹³. Studies of biodiversity all too often deal with species at the expense of interactions between those species, which are, after all, the driving force of ecosystem function. The more we know about such interactions, the better equipped we are to conserve them and to predict the outcomes of their disruption. □

Jeff Ollerton is at the School of Environmental Science, Nene-University College Northampton, Park Campus, Northampton NN2 7AL, UK.
e-mail: jeff.ollerton@nene.ac.uk

1. Proctor, M., Yeo, P. & Lack, A. *The Natural History of Pollination* (HarperCollins, London, 1996).
2. Pauw, A. *Nature* **394**, 731–732 (1998).
3. van der Pijl, L. *Evolution* **15**, 44–59 (1961).
4. Grant, V. *Proc. Natl Acad. Sci. USA* **91**, 10407–10411 (1994).
5. Waser, N. M., Chittka, L., Price, M. V., Williams, N. M. & Ollerton, J. *Ecology* **77**, 1043–1060 (1996).
6. van der Pijl, L. *Recueil Trav. Bot. Neerl.* **34**, 157–167 (1937).
7. Waser, N. M. in *Pollination Biology* (ed. Real, L. A.) 241–285 (Academic, New York, 1983).
8. Herrera, C. M. in *Floral Biology* (eds Lloyd, D. G. & Barrett, S. C. H.) 65–87 (Chapman & Hall, New York, 1996).
9. Ollerton, J. *J. Ecol.* **84**, 767–769 (1996).
10. Bruyns, P. & Linder, H. P. *Bot. Jahrb. Syst.* **112**, 453–527 (1991).
11. Ollerton, J. & Liede, S. *Biol. J. Linn. Soc.* **62**, 593–610 (1997).
12. Kearns, C. A. & Inouye, D. W. *BioScience* **47**, 297–307 (1997).
13. Tilman, D. et al. *Science* **277**, 1300–1302 (1997).

Daedalus

Defossilization

When a dead creature is fossilized, its organic remains are replaced by a mineral. Many fossils preserve the structure of the original organism with microscopic fidelity, as if the tissue had been replaced molecule by molecule by precipitated mineral — silica, calcium carbonate, or whatever. The dead organism is presumably permeated by a slightly supersaturated solution of the mineral. As each organic molecule is decomposed to soluble matter, its place is taken by deposited mineral. The process goes so extremely slowly that Daedalus sees it as an example of thermodynamic reversibility. Each organic molecule exchanges with an equivalent volume of mineral only when the process can go with no rise in entropy. The deposited mineral will adopt a crystal habit or microstructure whose strain-energy or grain-boundary density corresponds to, and therefore encodes perfectly, the free energy of the dissolving organic.

So Daedalus wants to reverse the process. Imagine, he says, a silica or carbonate fossil immersed in water quite free of mineral, but saturated with a wide range of biological substances: fats, sugars, proteins, DNA, and so on. The mineral will dissolve very slightly in the water. Its place will be taken by whatever biological molecules are best fitted to undertake the exchange deposition. These, of course, will be the ones originally replaced by the mineral. Keep the solution circulating round the fossil, remove the mineral from it as fast as it dissolves, top up the biochemicals as fast as they are deposited; and the whole process of fossilization will be gradually reversed. It may take years; but ultimately the original specimen will be exactly reconstituted.

This audacious process will require extreme biochemical subtlety. If the right substances are not present in solution, the defossilization mechanism will accept the closest available but incorrect match — especially if the process is hurried, thus losing reversibility. The next layer of organic, influenced by this mismatch, will be more mismatched still, and the reconstruction will deviate more and more from the original. Many fossils of the same type will have to be sacrificed upon the learning curve before a flawless specimen is finally recovered. But Daedalus dreams of the day when a perfectly reconstituted ancient frog or crab or fish opens its eyes in wonder, wiggles its body, and swims off into the solution.

David Jones

Obituary

James Lighthill (1924–98)

Applied mathematician and fluid dynamicist

Fluid dynamics developed and has continued to thrive through its applicability to fields of great practical importance. Dominant among these, during the first half of this century, was aerodynamics, which grew from an embryonic understanding of boundary-layer theory, flow instability, and subsonic and supersonic flow. Since about 1960, the dynamics of ocean and atmosphere, with its intricate interplay of wave-motion and turbulence, has emerged as a field of comparable vitality. And since about 1970, the fluid dynamics of biological systems has developed as a field of immense challenge now ripe for rapid growth.

James Lighthill's meteoric career was characterized by pioneering contributions in each of these areas. His work was based on classical techniques of applied mathematics, but his style involved strenuous efforts to interpret the mathematics, however complex, in terms of fundamental physics. It was this aim, so successfully achieved, that gave his papers their renowned lucidity and impact.

Michael James Lighthill was educated at Winchester College and Trinity College, Cambridge, where he read mathematics for the accelerated two-year wartime degree, graduating in 1943. His inclination was to undertake research in pure mathematics, but wartime imperatives dictated otherwise, and he moved directly to the National Physical Laboratory to work on supersonic aerodynamics. There he worked with ferocious intensity. Elected to a research fellowship at Trinity College in 1945, he was attracted in the following year to the University of Manchester by Sydney Goldstein, where he remained until 1959; he succeeded Goldstein as Beyer Professor of Applied Mathematics in 1950 at the exceptionally early age of 26.

By that time, the hazard of noise from jet aircraft was already recognized, and it was here that Lighthill made what many will consider his greatest contribution. In two papers, published in 1952 and 1954, he laid the foundations for all subsequent work on the subject. Recognizing that the turbulence in a high-speed jet is equivalent to a distribution of quadrupole sources, Lighthill was able to calculate the intensity (an eighth-power law) and directional distribution of the radiated sound field. An understanding of this mechanism of noise production was essential to the design of



jet engines whose noise output could be reduced to tolerable levels. Lighthill achieved worldwide recognition for this seminal work, and at the age of 29 was elected a fellow of the Royal Society.

From 1959 to 1964, Lighthill served as director of the Royal Aircraft Establishment at Farnborough, Hampshire. Despite the burdens of this post, he retained close contact with the wider research community, and indeed published papers in both geophysical and biological fluid dynamics, a foretaste of the interests that were to become his main preoccupation. These interests were given lease through his appointment in 1964 to a Royal Society professorship at Imperial College, London, and his subsequent appointment in 1969 as Lucasian Professor of Mathematics at Cambridge, a post he held for the next decade.

His great work during this period was concerned with waves in fluids (particularly fluids in rotation, as in the geophysical context) and also, most notably, with the subject that he named 'biofluidynamics'. Lighthill's interest in this topic had been stimulated much earlier by the Cambridge zoologist Sir James Gray; but it was only through his survey article, "Hydrodynamics of Aquatic Animal Propulsion", published in 1969, that the field really opened up. Lighthill himself emphasized the need for an interdisciplinary approach: "... if I was to help in classifying the hydrodynamics of aquatic animal locomotion I must talk to zoologists and go on talking to them; read their works and go on reading them; study their collections (in museums and aquaria)

and go on studying them!" And so he did. It was through this total immersion that a new realm of scientific endeavour was defined, explored and revealed to the fluid dynamics community.

In biofluidynamics, Lighthill contributed equally to our understanding of the flight of birds and of insects, topics for which his mastery of aerodynamics was well adapted. His appointment as provost of University College London (1979–89) did nothing to diminish his formidable research output. Again, his interaction with zoologists was of key importance. One such collaboration stands out — that with Torkel Weis-Fogh, a successor of James Gray at Cambridge — which led to elucidation of the mechanism of lift-production in small hovering insects. This is the clap-fling-sweep sequence by which circulation round each wing, and so lift, is generated by an essentially inviscid mechanism. Thus does the chalcid wasp *Encarsia formosa* achieve a lift coefficient greater than that of any man-made flying machine.

Lighthill was knighted in 1971, exactly halfway through his long research career. He was in great demand as a plenary lecturer at international meetings, and he would invariably rise to such occasions with gusto. His style was magisterial and he would present his material with total authority and little respect for any constraint of time. His lectures conveyed a passionate commitment to his subject, and rapier thrusts abound. Thus, for example, in 1962 when the magnetohydrodynamic bandwagon was in full swing: "It needs categorically to be reaffirmed that the continuum mechanics of a fluid innocent of electric current has as vital and exciting a present and future as any other branch of physical science". Lighthill's own work testified to the enduring truth of this statement.

Sir James Lighthill died on 17 July, at the age of 74, having almost completed a swim, in choppy seas, around the island of Sark in the English Channel, a feat that he had first accomplished 25 years earlier. He was a compulsive island circumnavigator, and his major source of pride in this respect was to have swum around Stromboli while it was erupting. Not for him the timid security of the seashore; he died as he had lived, with style and bravado. He is survived by his wife Nancy Dumaesq, to whom he was married for 53 years, and by one son and four daughters.

Keith Moffatt

Keith Moffatt is at the Isaac Newton Institute for Mathematical Sciences, 20 Clarkson Road, Cambridge CB3 0EH, UK.
e-mail: hkm2@newton.cam.ac.uk

Cartesian contrivances

If matter fills the Universe, making everything happen by its interactions, what does it all look like? René Descartes may have been over-mechanistic in his view, but his efforts to visualize the invisible created striking images.

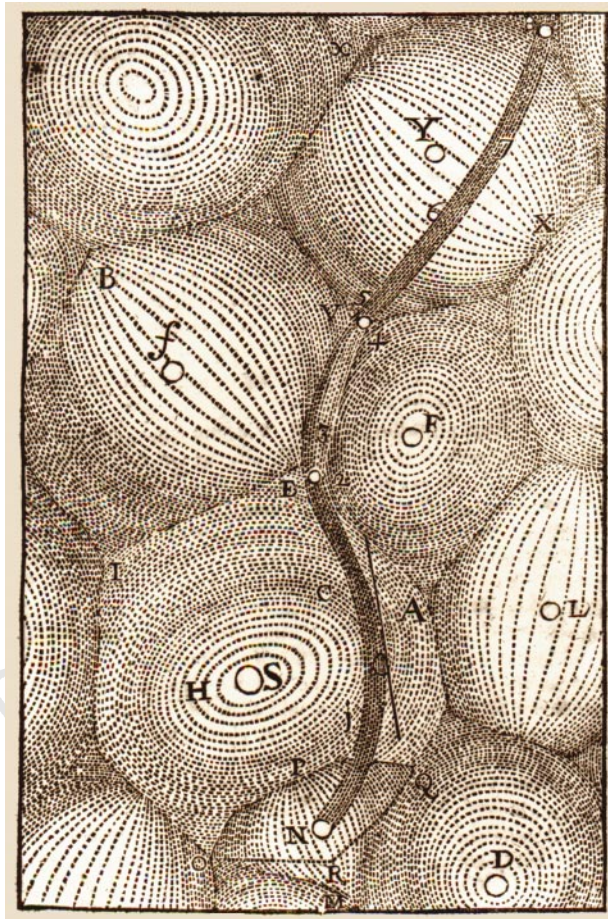
Martin Kemp

How to envisage and represent the ways that nature works has been the major challenge for illustrators of physical phenomena. Such things as invisible forces or the mechanisms responsible for 'action at a distance' can be described abstractly in words or encapsulated in mathematical formulas. But generations of natural philosophers and physicists have felt an apparently compelling need to develop and express their theories in more concrete terms, often by reference to existing machines or specially contrived mechanical analogies.

For a philosopher as obsessed with mechanism as René Descartes, the relationship between 'seeable' machines and the unseen machinery of God's cosmos was a matter of the greatest moment.

"The operations of things made by skill are, for the most part, performed by apparatus large enough to be easily perceived by the senses: for this is necessary so that they can be made by men. On the other hand, natural effects always depend on some devices so minute that they escape all senses."

To overcome this problem, Descartes' publications brilliantly exploit virtually all



Descartes' "celestial vortices", from *Principia Philosophiae*, 1644.

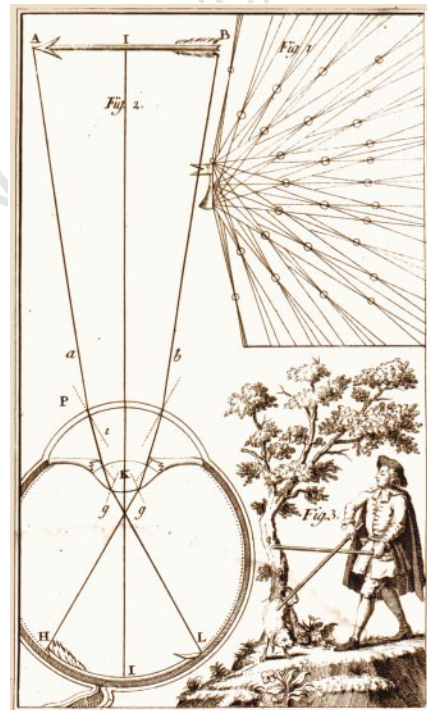


Illustration of the optics of the eye and Descartes' crossed sticks analogy, from *A Physical Essay on the Senses* by Claude Nicolas le Cat, 1750.

the modes of illustration available in the seventeenth century, ranging from pictorial representations to abstract diagrams.

He exhibited a particular genius for demonstrating invisible phenomena in terms of mechanical contrivances.

A number of his visual conceptions, such as his analysis of binocular vision and the inverted image on the retina in terms of a blind man with a pair of sticks, became oft-repeated classics.

Descartes was aware, however, that demonstration by mechanical analogy could be taken too literally — that the behaviour of the contrivance could all too easily be seen as constituting proof of what is actually happening with the phenomenon. Even for a philosopher who sought to explain all aspects of the physical world in terms of the action of matter on matter, illustration by mechanical analogy was a limited device.

He scolded those who viewed his aids to visualization in too obvious a way: "I did not say that light was extended like a stick, but like the actions or movements transmitted by a stick". The blind man probing with his

sticks simply helps us to understand how vision might work in terms of the underlying properties of things in action. The phenomenon and its mechanical visualization were analogous symptoms of the prime properties of magnitude, figure and motion.

His desire to convey the physical texture of even the most theoretical of his speculations is vividly shown in the 'picture' of his famous celestial vortices. The analogy here is hydraulic. Between the extremes of the most refined and luminous kinds of particulate matter and the coarse, weighty material of the Earth and planets were mobile globules that flowed in ceaseless whirlpools. These constituted a dynamic kind of cosmic foam that activated the apparent voids between celestial bodies. His extraordinary illustration of the Sun in the midst of its own vortex, packed within a three-dimensional system of contiguous vortices, stretches even his illustrative resources to breaking point. But it still does a vital job in a way that words could not emulate. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Pollen transfer on birds' tongues

Plants are adapted for pollination by birds in diverse and intricate ways. Here I describe an extraordinary example of these adaptations. The flowers of *Microlooma sagittatum*, a member of the milkweed family, clip parcels of pollen precisely onto the tongues of sunbirds. The pollen parcels are carried inside the bird's mouth to the next flower where they are detached mechanically. This is, to the best of my knowledge, the first evidence for pollen transfer on the tongues of birds, a phenomenon that might be more widespread.

Typical members of the cosmopolitan milkweed family (Asclepiadaceae) have their pollen contained in five removable parcels called pollinaria. Each pollinarium consists of two pollen-filled sacs and a mechanical clip, the corpusculum, for attachment to nectar-feeding insects. The corpusculum has a microscopic groove on one side, which catches on a bristle or other fine appendage on the body of an insect¹. All documented cases of pollination in the milkweed family are by insects².

Although it is a member of the milkweed family, the common South African climber *M. sagittatum* has many characteristics suggestive of bird pollination. Like many bird-pollinated flowers, and unlike typical insect-pollinated flowers³, the flowers of *M. sagittatum* are red in colour, secrete copious quantities of dilute nectar and are odourless at all times of day. The flowers also have the unusual characteristic of remaining in a bud-like state with the petals twisted together. The only access into the flower is through five tiny slit-like pores that insects would be unable to find or stretch open.

Despite observations of sunbirds visiting *Microlooma* flowers⁴, previous investigators^{1,4-7} assumed insect pollination in *M. sagittatum* and structurally similar flowers because birds seemed to lack a suitable site for the attachment of the minutely grooved corpusculum⁴. However, no specific pollinator has previously been identified for *M. sagittatum*.

I recorded visitors to the flower on 17 days and early evenings at three locations on the west coast of South Africa. No insects were seen on the flowers of *M. sagittatum*. Instead, lesser double-collared sunbirds (*Nectarinia chalybea*) were frequent visitors throughout the day (3.9 visits per plant per hour). A feeding bird probes the reinforced pores at the top of the flower with the sharp tip of its bill and extends its tongue into the bottom of the flower to reach the nectar (Fig. 1a).

By using flannel jackets to temporarily restrain mistnet-captured sunbirds⁸, I was

able to examine them under a dissecting microscope in the field, and discovered the pollinaria of *M. sagittatum* attached to the sunbirds' tongues (Fig. 1b). The tongue is tough and flexible, and rolled up lengthways from both sides to form twin capillary

tubes through which the nectar is drawn⁹. At its tip the tongue bifurcates and frays, and it is onto these frayed edges that the corpusculi become attached (Fig. 1c, d). Suitable sites for the attachment of the pollinaria can also be found on the tongues of



Figure 1 Mechanism of pollen attachment to bird tongues. **a**, A female lesser double-collared sunbird (*Nectarinia chalybea*) systematically sips nectar from the flowers of *Microlooma sagittatum*. Visits to a flowering plant lasted 21 ± 17 s ($n = 43$) with sunbirds probing 2.3 ± 0.6 flowers per s ($n = 21$). Plants bore 298 ± 219 flowers ($n = 8$). Flowers contained 0.61 ± 1.14 μ l of nectar at sunrise, which decreased to 0.07 ± 0.24 μ l after 6 h of sunbird activity ($n = 36$). When birds were excluded using gauze bags, nectar volumes reached 14.2 ± 6.8 μ l ($n = 19$). The nectar contained 23.6 ± 1.15 % sugar in sucrose equivalents ($n = 7$). Scale bar, 1 cm. **b**, The frayed tip of the tongue (white) of *N. chalybea*, with two attached pollinaria (yellow-brown) from the flowers of *M. sagittatum*. Each pollinarium consists of a corpusculum (brown) and two pollen sacs (yellow). Scale bar, 100 μ m. **c**, Detail of a pollinarium showing the groove in the corpusculum attached to the edge of the tongue. Scale bar, 20 μ m. **d**, Cross-section through the lower quarter of the flower. Outgrowths on the petals (o) guide the tongue down bristle-lined channels (b) that lead to the nectar. When the tongue is withdrawn, the bristles push it against the anthers (a) so that one of the frayed edges of the tongue is introduced into the guide rail (g). The guide rail leads into the narrowing groove of the corpusculum (above the level of sectioning), which attaches onto the tongue. The corpusculum, with two attached pollen sacs (p), is drawn up into the bird's mouth. When a bird visits the next flower with a pollinarium attached to its tongue, the row of bristles orientates the flattened pollen sac so it slides with its narrow edge into the guide rail, becomes wedged and breaks free from the clip as the tongue is withdrawn. Colours are near-natural; s, stigma; c, calyx. Scale bar, 250 μ m.

New World hummingbirds and Australian honeyeaters⁹.

Seven out of eight captured sunbirds were carrying pollinaria or partial pollinaria on their tongues, with up to four corpusculi attached at a time (median number of corpusculi per bird, 2; median number of pollen sacs per bird, 0.5). The frequent presence of corpusculi with only one or neither of their two pollen sacs attached suggests that pollen sacs were transferred successfully to other flowers.

As expected from the high levels of flower visitation, a large proportion (54%) of flowers were found to be pollinated ($n = 48$). Three times more pollen sacs were removed (2.90 ± 2.74 per flower) than re-inserted (0.96 ± 1.01 ; $n = 48$), indicating substantial losses during transfer. Seed pod production is high in plants exposed to sunbirds, whereas bagged (pollinator excluded) flowers set no seeds ($n = 119$).

A captive sunbird was used to test whether sunbirds successfully transfer pollen sacs to the intricate pollen-receiving mechanism (Fig. 1d). I exposed 37 virgin *M. sagittatum* flowers to the bird for a period of two minutes per flower. There were no insects in the cage. Dissection showed that the 37 experimental flowers received 39 pollen sacs, indicating that sunbirds effectively both collect and deliver pollen sacs.

Plants adapted for pollen transfer on bird tongues are characterized by having: (1) tough and streamlined pollinaria with a mechanical clip; (2) morphology that excludes insects; (3) reddish coloration; (4) copious, dilute nectar; and (5) no scent. Applying these criteria to the rest of the genus *Microlooma*⁵, I predict that six of the nine remaining species are pollinated in the manner described here. Outside Africa, many members of the large southeast Asian genus *Dischidia*¹⁰ are among other plants that may be adapted for pollen transfer on the tongues of birds. Anecdotal accounts of sunbirds, hummingbirds and honeyeaters visiting the flowers of several genera of Asclepiadaceae² beg investigation.

Anton Pauw

Department of Botany,
University of Cape Town,
Private Bag, Rondebosch 7701,
Cape Town, South Africa
e-mail: pauw@botzoo.uct.ac.za

- Kunze, H. *Plant Syst. Evol.* **176**, 227–253 (1991).
- Ollerton, J. & Liede, S. *Biol. J. Linn. Soc.* **62**, 593–610 (1997).
- Faegri, K. & van der Pijl, L. *The Principles of Pollination Ecology*. (Pergamon, Oxford, 1979).
- Vogel, S. *Bot. Stud.* **1**, 1–338 (1954).
- Bruyns, P. & Linder, H. P. *Bot. Jahrb. Syst.* **112**, 453–527 (1991).
- Wanntorp, H. E. *Opera Bot.* **98**, 1–69 (1988).
- Marloth, R. *The Flora of South Africa*, Vol. 3 (Darter, Cape Town, 1932).
- Kearns, C. A. & Inouye, D. W. *Techniques for Pollination Biologists* (Univ. Press of Colorado, Niwot, 1993).
- Paton, D. C. & Collins, B. G. *Aust. J. Ecol.* **14**, 473–506 (1989).
- Rintz, R. E. *Blumea* **26**, 81–126 (1980).

A proxy index of ENSO teleconnections

Variations in central and eastern equatorial sea surface temperatures are linked to extra-tropical climate changes by atmospheric heat and moisture fluxes in so-called 'teleconnection' patterns^{1,2}. Estimates of sea surface temperature (SST) anomaly may be derived from the isotopic composition and minor-element chemistry of aragonite formed by annually banded reef corals. Here we reconstruct anomalies in central equatorial Pacific SSTs for 1938–93 from the stable oxygen isotope ($\delta^{18}\text{O}$) composition of a coral collected at Kiritimati (Christmas) Island, Republic of Kiribati (157.3°W , 2°N). We recover not only the regional anomaly in SST, but also extra-tropical climate information in climatically teleconnected regions.

Interannual variability in the $\delta^{18}\text{O}$ composition of aragonite is a function of the temperature of the sea surface in which the coral secretes its skeleton³. This variability in coral $\delta^{18}\text{O}$ also records any changes in the $\delta^{18}\text{O}$ of sea water, which may be significant in some oceanographic settings⁴.

The strontium-to-calcium (Sr/Ca) ratio in corals is also a function of SST, but seawater Sr/Ca variability is minimal^{5,6}. Application of these two proxies in tandem has been used to estimate both seawater $\delta^{18}\text{O}$ and SST anomalies during the 1982–83 El Niño–Southern Oscillation (ENSO) event from measurements made on Great Barrier Reef corals⁷. In principle, paired Sr/Ca and $\delta^{18}\text{O}$ data from a series of temporally overlapping coral cores may yield proxy estimates of SST and seawater $\delta^{18}\text{O}$ variability extending well into the pre-instrumental period.

At Kiritimati, seasonal SST varies by $\leq 1.2^\circ\text{C}$, whereas ENSO-induced SST anomalies of $\pm 3^\circ\text{C}$ reflect changes in wind stress and associated equatorial upwelling⁸. Thus, Kiritimati is an ideal location at which to monitor the thermal oceanographic signal associated with the full ENSO cycle. We measured $\delta^{18}\text{O}$ and carbon isotopic ($\delta^{13}\text{C}$) composition in a core from a living coral, collected in March 1994, at intervals of 0.5 mm along the maximal growth axis of the coral.

We used the $\delta^{13}\text{C}$ record to construct an age model for analysis of the $\delta^{18}\text{O}$ record over time, on the basis of observations of the dependence of $\delta^{13}\text{C}$ levels in coralline aragonite on coralline reproductive

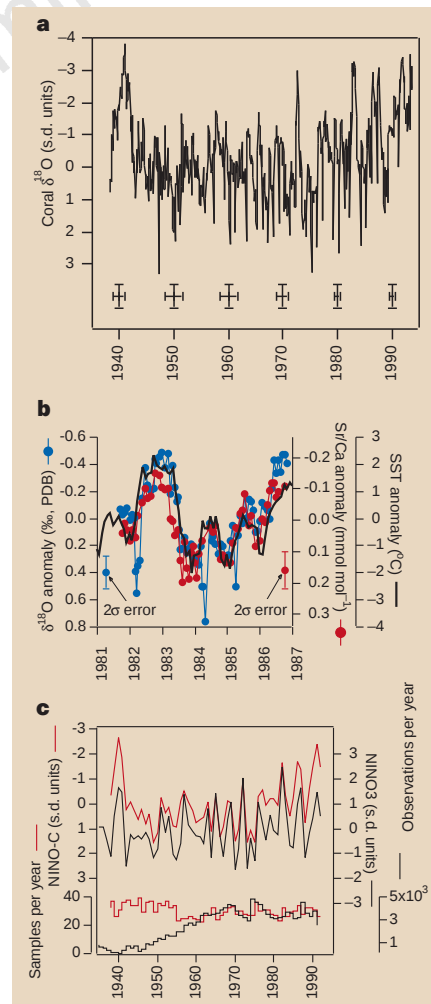


Figure 1 Kiritimati coral proxy data. **a**, Kiritimati coral $\delta^{18}\text{O}$ time series normalized to the 1951–80 mean and standard deviation of the data. The $\delta^{18}\text{O}$ scale has been reversed for direct comparison with SST anomalies (**b**). Vertical error bars give uncertainty in $\delta^{18}\text{O}$ measurements (2σ) and horizontal error bars give decadal averages of age-model uncertainty in units of year/age-model year¹⁰. s.d. units, standard deviation units. **b**, Kiritimati coral $\delta^{18}\text{O}$ anomalies (‰ relative to the Pee Dee Belemnite (PDB) isotopic standard) and Sr/Ca anomalies (mmol mol⁻¹) relative to the mean for the 1981–87 period. Reproducibility confidence intervals (estimated 2σ) for each tracer are as indicated. Also plotted are estimates of SST anomalies from the analysis by Reynolds and Smith¹⁴ for the $2^\circ \times 2^\circ$ grid box encompassing Kiritimati. The data have been plotted on a common range so that direct comparisons of the different measurements may be made: $8^\circ\text{C} \approx 1.6\text{‰}$ in $\delta^{18}\text{O} \approx 0.6 \text{ mmol mol}^{-1}$ in Sr/Ca^{3,6}. **c**, Plot of NINO-C and NINO3 indices (SST anomalies averaged over $150^\circ\text{--}90^\circ\text{W}$, 5°N to 5°S) against time. Also shown are the numbers of samples analysed per year and of SST observations per year. Correlation between NINO-C and NINO3 is -0.79 .

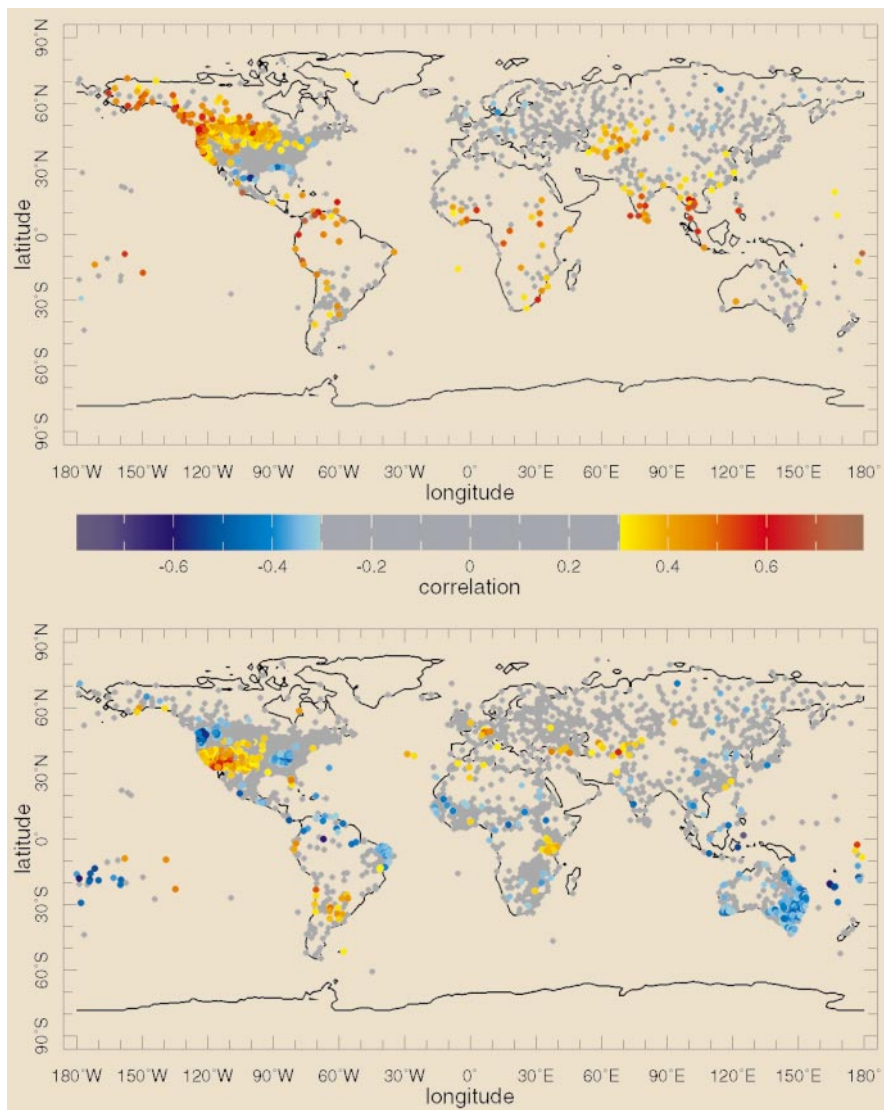


Figure 2 Linear correlations (r) between NINO-C and surface air temperature (top) or precipitation data (bottom) from a global network of stations¹⁵. The correlation maps have been multiplied by -1 so that positive (negative) correlations are associated with positive (negative) SST anomalies. The locations of all stations with ≤ 42 years' worth of data are plotted in grey; stations correlated at or above the 90% confidence level ($|r| \geq 0.3$; ≈ 36 degrees of freedom) are plotted in colour. In general the recovered patterns agree with those recovered by correlation with like averages of NINO3, although there are subtle differences.

activity⁹. This sampling resolution allowed us to obtain 25–35 samples per annual band and to achieve a chronology with better-than-monthly resolution and low age-model error estimates¹⁰ (Fig. 1a).

Paired $\delta^{18}\text{O}$ and Sr/Ca results and SST data for the 1981–86 period confirm that both geochemical thermometers captured the interannual SST variability over this six-year period, which includes the strong ENSO warm event of 1982–83, normal/cold conditions throughout mid-1986, and the initiation of the 1986–87 warm-phase event (Fig. 1b). As our interest here is in the interannual variability, we analysed further only April-to-March annual averages of the oxygen isotopic record, which we term 'NINO-C' and interpret as being primarily driven by SST anomalies (Fig. 1c). NINO-C shares over 60% variance with the commonly used

NINO3 area index (150°–90° W, 5° N to 5° S) of the eastern equatorial Pacific SST anomaly¹⁰.

NINO-C should also provide a good index of teleconnections associated with ENSO, as SST extremes in the central equatorial Pacific directly influence the atmospheric circulation^{1,2}. Analysis of correlation patterns between data from surface meteorological stations and NINO-C indicates that global, interannual patterns of surface air temperature and precipitation anomalies are recoverable from NINO-C (Fig. 2) in the Americas, Africa, Australia, India and southeast Asia.

These patterns generally agree, in both spatial distribution and correlation amplitude, with those recovered by correlation of station data with like averages of NINO3 and with the literature on tropical–

extratropical climate teleconnections^{1,2,11–13}. However, correlations with NINO3 are stronger than correlations with NINO-C in the southeast United States for both temperature and precipitation, indicating that this region may be more sensitive to SST variability in the eastern than the central equatorial Pacific.

In contrast, surface temperatures in central and northeastern Asia at about 45° N appear to be responsive to NINO-C but not to NINO3. Thus there appear to be subtle differences between teleconnections associated with the central and eastern equatorial Pacific regions as well. These differences may be resolved by proxy indices from the central equatorial Pacific and other data-poor regions of the tropical oceans.

We have shown that NINO-C can provide a record of both tropical and extratropical climate variability on interannual time scales. Thus, comparison of modern and palaeoclimatic NINO-C results will allow analysis of climate change relative to the present, both in the central equatorial Pacific and in teleconnected regions. Seasonally variable carbon isotopic composition or density banding provides a relative chronology for a coral colony's lifetime, and absolute dating of fossil coral samples can be achieved by the radiocarbon or U-series techniques. By these means, multiproxy extension of NINO-C into the pre-instrumental period will allow exploration of links to documented extratropical palaeoclimate anomalies on interdecadal time scales.

M. N. Evans*†, R. G. Fairbanks*†, J. L. Rubenstone*

**Lamont–Doherty Earth Observatory of Columbia University, Palisades, New York 10964-8000, USA*

†*Department of Earth and Environmental Sciences, Columbia University, New York, New York 10027, USA*

e-mail: mevans@ldeo.columbia.edu

1. Pan, Y. H. & Oort, A. H. *Mon. Weath. Rev.* **111**, 1244–1258 (1983).
2. Graham, N. E. & Barnett, T. P. *Science* **238**, 657–659 (1987).
3. Weber, J. N. & Woodhead, P. M. J. *J. Geophys. Res.* **77**, 463–473 (1972).
4. Cole, J. E., Fairbanks, R. G. & Shen, G. T. *Science* **260**, 1790–1793 (1993).
5. Weber, J. N. *Geochim. Cosmochim. Acta* **37**, 2173–2190 (1973).
6. DeVilliers, S., Shen, G. T. & Nelson, B. K. *Geochim. Cosmochim. Acta* **58**, 197–208 (1994).
7. McCulloch, M. T., Gagan, M. K., Mortimer, G. E., Chivas, A. R. & Isdale, P. J. *Geochim. Cosmochim. Acta* **58**, 2747–2754 (1994).
8. Philander, S. G. H. *El Niño, La Niña, and the Southern Oscillation* (Academic, San Diego, 1990).
9. Gagan, M. K., Chivas, A. R. & Isdale, P. J. *Geology* **24**, 1009–1012 (1996).
10. Evans, M. N. *et al.* *J. Geophys. Res.* (submitted); preprint available at <http://rainbow.ldeo.columbia.edu/~mevans/preprints/> (1998).
11. Ropelewski, C. F. & Halpert, M. S. *Mon. Weath. Rev.* **114**, 2352–2362 (1987).
12. Ropelewski, C. F. & Halpert, M. S. *J. Clim.* **2**, 268–284 (1989).
13. Halpert, M. S. & Ropelewski, C. F. *J. Clim.* **5**, 577–593 (1992).
14. Reynolds, R. W. & Smith, T. M. *J. Clim.* **7**, 929–948 (1994).
15. Baker, C. B. *et al.* in *Preprints of AMS Ninth Conf. Appl. Climatol.*, Dallas, 15–20 Jan. 1995; also <http://ingrid.ldeo.columbia.edu/SOURCES/NOAA/NCDC/GCPS/> (1998).

Trifling variation in truffles

Of the ten species of European truffle (fungi of the genus *Tuber*, phylum Ascomycota), some have economic value because of their organoleptic properties (taste and perfume), in particular the black truffle (*Tuber melanosporum* Vitt.) and the summer and burgundy truffles^{1,2}. The black truffle is mainly found in Spain, France and Italy (Fig. 1a), and it shows variation in several traits, including in its famous organoleptic properties, across this geographical range. Here we show that this variation probably results from environmental, rather than genetic, influences.

In an attempt to explain the variation in *T. melanosporum* across its geographical range and to study the distribution of genetic variability within and among populations, we analysed fruiting bodies (ascocarps) from different populations in France and Italy for random amplified polymorphic DNA (RAPD) and microsatellite polymorphism.

We found an extremely low level of polymorphism over the whole study area for both types of marker (Fig. 1b). This pattern contrasts dramatically with that of the sympatric summer truffle (*T. aestivum* Vitt., which forms a species complex with the burgundy truffle, *T. uncinatum* Chatin), which has distinct internal transcribed spacer alleles³ and highly variable RAPD patterns⁴.

The absence of heterozygotes in *T. melanosporum*, shown by study of microsatellites, is consistent with a very closed mating system, such as homothallism or even exclusive selfing. As the two putative haploid genomes from an ascocarp are very similar, we considered each variable RAPD band to be a distinct locus exhibiting two alleles (presence or absence of a band). The genetic models used subsequently showed no increase in genetic differentiation with geographic distance (Fig. 1c).

The low level of genetic diversity of the black truffle probably cannot be explained by its present large population size. In France, 1–3 × 10⁴ kilograms of black truffles are now officially sold each year, and more than 10⁶ kilograms per year were traded in the last century (the average weight of an

ascocarp is about 30–40 grams).

A population bottleneck probably occurred during the last, and coldest, glaciation, when the broadleaved forest of Europe was considerably reduced and restricted mainly to the Mediterranean coastal zone⁵. The black truffle ripens in winter (November–February), which probably contributed to its drastic reduction in population size and restriction to its southernmost limits during the glaciation, as ascocarps of contemporary *T. melanosporum* are susceptible to frost. The present low level of genetic variability in black truffle populations is consistent with such a bottleneck occurring 10,000 years ago, followed by a rapid colonization of southwestern Europe, which would also explain the absence of phylogeographic signals in the few polymorphic markers found. The 'glaciation hypothesis' would be strengthened if more southerly populations (in Spain or Italy) were found to show greater genetic diversity.

The seasonal behaviour of the summer and burgundy truffles, which ripen in spring and autumn, respectively, would have allowed them to sustain a larger geographical range and population size during the last glaciation, explaining the present high level of genetic variability of this species complex. Moreover, their current geographical range, extending further to the east and north (for example, to Poland and Sweden⁶), shows that they are more tolerant of colder climates.

Our results show that the morphological and organoleptic differences seen over the geographical range of the black truffles can probably be explained by environmental variation rather than by genetic factors. Research is needed to identify the environmental variables that affect the black truffle's perfume and taste, which are the objects of intense human interest.

G. Bertault, M. Raymond, A. Berthomieu
Institut des Sciences de l'évolution (UMR 5554),
Laboratoire Génétique et Environnement, Cc 065,
Université Montpellier II, 34095 Montpellier,
Cedex 05, France.

e-mail: raymond@isem.univ-montp2.fr

G. Callot

UFR Sciences du Sol,
Institut National de la Recherche Agronomique,
Place Viala, 34060 Montpellier, Cedex 01, France

D. Fernandez

Laboratoire de Phytopathologie Tropicale,
ORSTOM, BP 5045, 34032 Montpellier,
Cedex 01, France

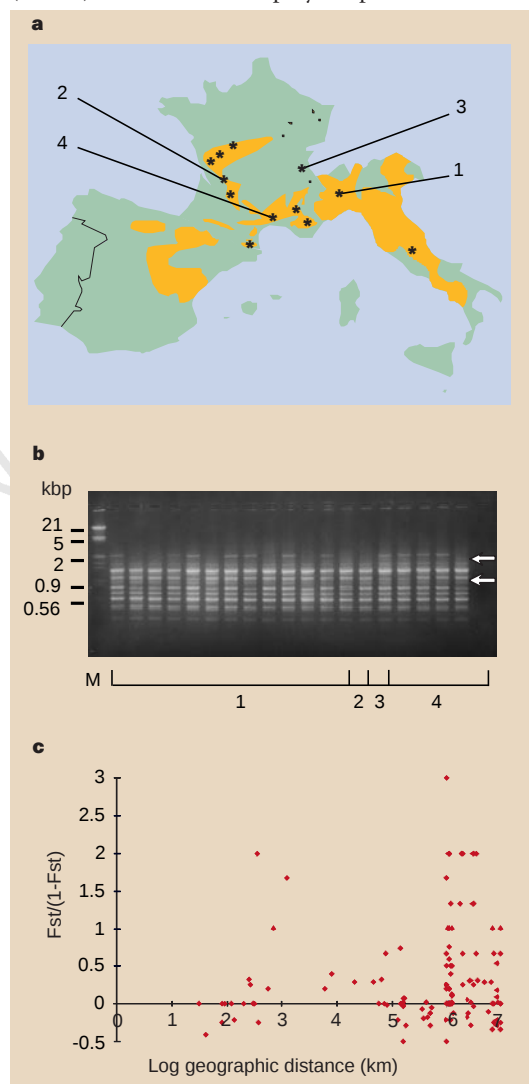


Figure 1 Geographic and genetic characteristics of the black truffle *Tuber melanosporum* Vitt. **a**, The western European geographical range (shaded area) of *Tuber melanosporum* and sampling localities (stars); numbers refer to samples shown in **b**. All truffles ($n=208$) were collected in natural habitats, except one sample which was collected in an artificial 'truffle-field'. **b**, Example of RAPD patterns obtained with the OPF-14 primer (Operon Technologies, Alameda) for truffles from four locations (numbers refer to **a**); arrows indicate polymorphic bands used. Sizes on left are in kilobase pairs. **c**, Analysis of isolation by distance according to ref. 7, using six RAPD loci (generated by primers OPF-11, OPF-13, OPF-14 and OPB-2) and the only two polymorphic microsatellite loci out of the nine assessed, (GAGT)₁₄ and (GTTA)₉, showing two alleles each. Computations were performed with Genepop version 3.1b (ref. 8). The increase in genetic differentiation with geographic distance was not significant (Mantel test, 10⁵ permutations, $P>0.10$).

1. Henrion, B., Chevalier, G. & Martin, F. *Mycol. Res.* **98**, 37–43 (1994).
2. Amicucci, A. et al. *Biotechnol. Lett.* **18**, 821–826 (1996).
3. Guillemaud, T. et al. *Mycol. Res.* **100**, 547–550 (1996).
4. Gandeboeuf, D. et al. *Can. J. Bot.* **75**, 36–45 (1997).
5. Blondel, J. *Biogéographie, Approche écologique et évolutive* 1–297 (Masson, Paris, 1995).
6. Chevalier, G. & Frochet, H. *La Truffe de Bourgogne (Tuber uncinatum Chatin)* 1–257 (Pétraque, Levallois-Perret, 1997).
7. Rousset, F. *Genetics* **145**, 1219–1228 (1997).
8. Raymond, M. & Rousset, F. *J. Hered.* **86**, 248–249 (1995).

Enigma variations on the nuclear stage

Copenhagen

A play by Michael Frayn
Performed at the National Theatre, London
Methuen: 1998. 128pp. £6.99 (pbk)

Michael Berry

In *Copenhagen*, Michael Frayn re-creates one of the more mysterious episodes of the Second World War, when Werner Heisenberg travelled to German-occupied Denmark in 1941 to visit his fellow physicist Niels Bohr. Why did he go there?

Historians — as well as the protagonists themselves, with their faulty and shifting memories — have long puzzled over Heisenberg's motives. Did he seek to enlist Bohr's assistance in developing a German nuclear bomb? Was he pumping Bohr for information about a possible parallel programme by the Allies? Was he alerting Bohr to the existence of the German programme, in the hope that Bohr would provoke the Allies to accelerate their own programme? Was he seeking a reason to delay, or even sabotage, the German nuclear effort? And why did the German programme fail? Was it through rivalry and division of resources between competing teams, or because a crucial neutron diffusion rate was wrongly assumed instead of being calculated?

In the play, these possibilities are explored in detail and with sensitivity, concentrating on the contrapuntal relationship between the two physicists' personalities, their careers, and their scientific styles. Bohr, 16 years Heisenberg's senior, initiated the modern theory of the atom in 1913, with his proposal that the energies of electrons in atoms are restricted by quantum rules similar to those that Planck and Einstein had applied to light. This was not only bold but outrageous, because it had no basis in theoretical physics — indeed, it flatly contradicted the established 'classical' physics of the day.

Twelve astonishing years followed, in which the most intense concentration of scientific effort was devoted to the search for the fundamental quantum theory underlying Bohr's atom. An important centre of that research was Bohr's newly established institute in Copenhagen. He imported a stream of brilliant young physicists, and sought unremittingly to uncover the physical significance of their theories and formalisms. This led him into philosophy, and to his principle of complementarity. His style was slow, tentative, almost inarticulate, repeatedly redrafting his papers in the hope of reaching a version that matched what he was groping for.

Heisenberg, competitive, insecure, barely into his twenties when he went to Copenhagen, was instantaneous in his responses, diamond-hard in the precision of his



Meeting of minds: Burke (left), Kestelman and Marsh succeed in the fusion of science and drama.

thought. He occupied himself with mathematical formalism rather than wordy interpretations. It was he who reached the first consistent quantum mechanics: spare, algebraic, a landscape alien to theoretical physicists yet capable of reproducing all the hitherto baffling experimental observations on atoms. Complementing Bohr's complementarity was Heisenberg's uncertainty principle, a precise statement of the limits within which a particle's position and motion can be known.

The structure of the play mirrors these contrasts. Bohr appears mumbling, reflective, but with a simple goodness ("I don't think anyone has yet discovered a way you can use theoretical physics to kill people") and clear moral insight; Heisenberg, swift-tongued but morally ambiguous. The 1941 meeting is repeatedly re-enacted and revisited, echoing Bohr's repeated redraftings, in attempts to get at what really happened. As each dilemma gets focused on, another recedes, vague, into a mist.

Underlying the momentous events that followed the Copenhagen encounter are facts about atomic nuclei, understandable only in terms of quantum physics. Frayn does a splendid job of explaining these subtle and tricky matters, in some detail yet without technicalities. Through the exchanges of the protagonists, we get clear accounts of fission, the production of nuclei, chain reactions (where the play as staged corrects an arithmetic error in the published version), the important distinction between slow and fast neutrons, diffusion rates, and quantum uncertainty and interference.

On a stage bare but for three chairs, our attention is gripped and held by three actors: Bohr, played by David Burke, Heisenberg by Matthew Marsh, and Bohr's wife Margrethe by Sara Kestelman. Margrethe is the chorus,

mercilessly exposing Heisenberg's evasions and Bohr's good-natured, even fatherly, indulgence towards his former colleague. Marsh and Burke bring out beautifully the obvious affection between the two very different men, strained to the utmost by the excruciatingly hard times — and not helped by Heisenberg's clumsy invitation to the Bohrs to make use of his ski-hut in Bavaria, apparently forgetting that Bohr was half-Jewish.

With *Copenhagen*, Frayn helps to create a genre, alongside Tom Stoppard, with *Hapgood*, inspired by quantum mechanics, and *Arcadia*, based on chaology, and Mike Maran, with *Surely You're Joking, Mr Feynman*, about that unique and colourful physicist. This acceptance of science as a legitimate subject for dramatization, rather than something separate and technical, is both welcome and overdue. In human culture, as with nuclei, fusion is more powerful than fission but harder to achieve. □

Michael Berry is in the Physics Department, University of Bristol, Bristol BS8 1TL, UK.

Myths that won't die

Frankenstein's Footsteps: Science, Genetics and Popular Culture

by Jon Turney

Yale University Press: 1998. 320pp. £19.95, \$30

Roslynn Haynes

"What terrified me will terrify others", wrote Mary Shelley of the nightmare that allegedly provided the inspiration for *Frankenstein*. Nearly two centuries later her words remain uncomfortably prescient. The name of her protagonist and the image of his Monster are ever available as shorthand to evoke the whole package of emotional reactions to perceived hubris — from guilty fascination to fear and

moral revulsion — with which Western culture periodically frightens itself.

The locus of our concern may have moved from the electrical stimulation of corpses, through vivisection and eugenics, to the Human Genome Project, cloning and experiments with recombinant DNA, but we have failed to come up with new scientific myths in which to express it.

Jon Turney's comprehensive study of the public face of biology since Shelley's time shows how her creation has never let go its power, even stranglehold, over popular imagination. The literary inheritance of *Frankenstein* has been well documented before. But Turney's important contribution lies, first, in integrating the fictional images with their counterparts in film and journalism and, second, in attempting to assess the appropriateness of the myth today, in relation to what Dorothy Nelkin and Susan Lindgee have called the "narratives of genetic essentialism".

Formerly science editor of the *Times Higher Education Supplement* and currently senior lecturer in science communication at University College London, Turney is admirably qualified to trace journalists' responses to the intellectual and technological leaps biologists have made in this century. Drawing on sources as diverse as reports in *Nature*, tabloid headlines and cartoons, he demonstrates how consistently *Frankenstein*, and the related mad-scientist script, have been invoked. In the process he also presents us with an extremely readable, if selective, history of the discipline and frequently points to the disparity between popular impressions of biological landmarks and the views of the scientists concerned.

On the one hand, Turney shows the inadequacy of the stereotyped response to technological breakthroughs. The furore greeting the birth of Louise Brown, the world's first 'test-tube' baby, now seems astonishing, generating fears that have proved groundless. Perhaps, in a decade, the doomsaying response to Dolly, the cloned sheep, will appear equally overblown. But on the other hand, the bland dismissal by many scientists of all social concerns as mere anti-scientism, as 'genetic pornography', is equally blinkered.

More sinister than the facile name-calling is the cynical manipulation of public anxiety. A notable case documented here was the moratorium declared in 1974 by molecular biologists on their own recombinant DNA experiments. By this seemingly responsible initiative, the signatories not only softened public opinion but retained control over both the duration of the moratorium and the boundaries of the ensuing discussion. The focus was effectively and conveniently limited to technological safety measures rather than embracing issues of social and moral responsibility that were the real crux of popular concern.



No fear: with hindsight, public anxiety about biological breakthroughs may seem unfounded.

The acknowledged inspiration for *Frankenstein's Footsteps* was *Nuclear Fear* (Harvard University Press, 1988), Spencer Weart's history of images that attempted to uncover the origins of our attitudes to nuclear physics, and a particularly interesting aspect of Turney's study is the juxtaposition of popular responses to biology and physics. During the Cold War, atomic physicists were designated the heirs of Frankenstein, but Turney argues that biology remains the more dangerous, more invasive discipline.

Where physics is largely concerned with understanding and with theory, biology openly aligns itself with the Baconian aspiration of manipulating knowledge to "the effecting of all things possible". Indeed, biologists have frequently been the ones to introduce provocative rhetoric, from Jacques Loeb and Alexis Carrel's self-publicity at the beginning of the century about their imminent 'creation of life' in the laboratory, to Francis Crick's emotive phrase for molecular biology as an exploration of "the borderline between the living and the dead", with its echo of Frankenstein's words. Michael Mulkay's recent analysis of the British parliamentary debate over embryo research shows that, while non-scientists rarely invoked Frankenstein or his successors, scientists themselves introduced such references to the record by claiming that their opponents were unduly influenced by them.

Although he strives to be factual and unbiased throughout, Turney seems to suggest

that many of the fears voiced by the 'uninformed' public are not entirely unfounded. For a century biologists have a sad record of treating society with contempt. His conclusion is that, rather than accusing their lay contemporaries of being fixated on the Frankenstein cliché, scientists themselves need to be more innovative and serious in telling new and more complex stories. These could help us move beyond the false, polarizing alternatives of closed minds and created monsters.

Although this book necessarily concentrates on written texts, the black and white illustrations form an important adjunct, indicating the extent and flexibility of the imagery surrounding the subject. For example, a recent cartoon from *The Guardian* shows a figure who not only bares his manly chest but parts the flesh like a model in a classical anatomy text, to reveal a bar code representing his DNA analysis — the ultimate identity index. *Frankenstein's Footsteps* should be read by all who are concerned about the social and ethical implications of DNA experimentation and, even more importantly, by those who are not. □

Roslynn D. Haynes is the author of *From Faust to Strangelove: Representations of the Scientist in Western Literature* (John Hopkins University Press, 1994). She is in the School of English, University of New South Wales, Sydney, NSW 2052, Australia.

Getting to grips with the solid Earth

Principles of Geophysics

by Norman H. Sleep and Kazuya Fujita
Blackwell Science: 1997. 586pp. £39.50, \$72.95

Fundamentals of Geophysics

by William Lowrie
Cambridge University Press: 1997. 354pp.
£55, \$85 (hbk); £19.95, \$34.95 (pbk)

Jon Bull

Geophysics is a broad term that encompasses all aspects of the Earth system. These textbooks focus on "solid earth geophysics" but also have chapters on planetology. Both aim to be comprehensive supporting texts for university courses; they are similar in scope and range to the widely used *The Solid Earth* by C. M. R. Fowler (Cambridge University Press, second edition due summer 1999).

Principles of Geophysics, by Norman H. Sleep and Kazuya Fujita, is well written, adopts a novel approach, and concentrates on the underlying physics. The level of difficulty increases through the book, but each chapter has a clearly written introduction in which the key physical concepts are outlined. Most aspects of solid Earth geophysics are dealt with, but in covering such a vast range of material the treatment is necessarily

brief. The mathematical derivations are well explained, but for the average student most chapters would need to be studied in close conjunction with other sources.

The content is original (for example, very few diagrams are reproduced from other sources), the presentation of the book is good, and there seem to be few errors or inaccuracies. A feature that works particularly well is the selected reading list at the end of each chapter in which key references are given and their content summarized. The level is generally more advanced than in Fowler, but is equally well explained.

In *Fundamentals of Geophysics*, William Lowrie hopes to provide a core textbook for geology and geophysics undergraduates. But aiming a book at this intermediate level is difficult: you need to ensure that material on each topic is adequately introduced and also that the treatment does not frustrate the reader with its lack of depth. Lowrie's text is certainly not advanced, but his introduction to material is frequently inadequate. The level is similar to Fowler, but less clearly explained and presented.

While it thoroughly covers solid Earth geophysics, some aspects of the book are ill-conceived. For example, the chapter on seismology starts by stating that it is necessary to have a good grasp of elasticity theory before analysing the different sorts of seismic wave. I would not agree with this. Lowrie presents the full derivation of the wave equation before more qualitative aspects, such as the seismograph, seismogram and earthquake seismology, but this does not work well. Sleep and Fujita neatly split their treatment of seismology into an earlier qualitative chapter and a later more quantitative one. Fowler's solution is to present the wave equation as an appendix, which is also a more satisfactory approach.

Another problem with Lowrie is that the level of mathematical difficulty is very inconsistent. Also, he generally introduces a topic from a historical viewpoint. While this is interesting, I think that it is more important to start with the underlying physics and then present the geophysical applications (perhaps with historical references), which is essentially the approach followed by Sleep and Fujita.

Neither of these new books has problems or exercises for the student to reinforce the contents of each chapter, which is a pity. I have found the problem sections in Fowler's text particularly useful, not only as set for students, but also for adaptation as examples in class lectures.

In summary, I wholeheartedly recommend *Principles of Geophysics* to any research geophysicist or graduate student, and as a key library text for teaching. I will continue to recommend Fowler as the course text, but will refer students to Sleep and Fujita for particular aspects (the chapter on heat flow and

geothermics, for example, is outstanding). Lowrie's text needs substantial revision before I could recommend it for the support of teaching. □

Jon Bull is in the School of Ocean and Earth Science, Southampton Oceanography Centre, University of Southampton, Southampton SO14 3ZH, UK.

Defining moment

The Environment Dictionary

by David D. Kemp

Routledge: 1998. 464pp. \$100 (hbk, not yet published); £17.99, \$29.99 (pbk)

Longman Dictionary of Environmental Science

by Eleanor Lawrence, Andrew R. W.

Jackson and Julie M. Jackson

Longman: 1998. 491pp. £12.99 (pbk)

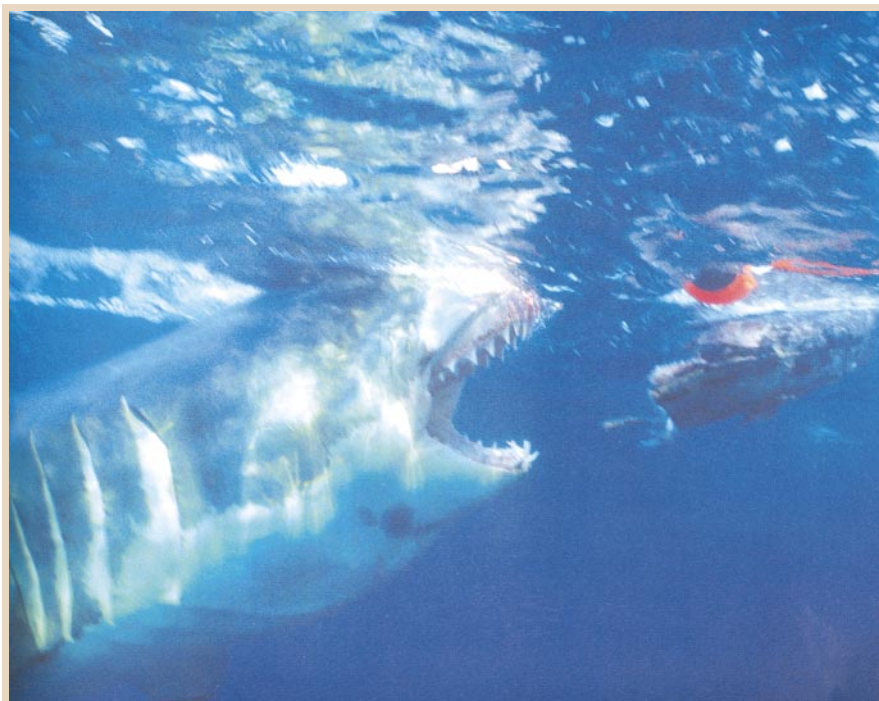
Peter Brimblecombe

The maturity of a subject can often be gauged by the development of its specialist reference works: dictionaries, encyclopedias and handbooks. Many traditional disciplines offer fine examples of these, but the growing importance of the environmental sciences is only beginning to be reflected in the quality of its reference works. The environment is of such widespread interest that it has created the need for both popular and specialist

reference books. These range from handbooks for active researchers through to volumes that help ordinary people confronted by the barrage of environmental terminology now in everyday use.

Although both David Kemp's *The Environment Dictionary* and the *Longman Dictionary of Environmental Science* aim at fairly general readers, they are very different. The former provides 1,700 general entries in the style of an encyclopedia. These are often some pages in length and well supplied with advice on further reading. The latter offers nearly 13,000 short definitions.

What is environmental science? This has always been a difficult question, so it provides an appropriate first test for these two books. Under "environment" Kemp tells us that "environmental science includes not only the traditional sciences such as chemistry, physics and biology, but also engineering, economics, sociology, politics and law. The study of the environment is thus very much interdisciplinary in nature". The Longman dictionary defines it as "the study of how humans and other species interact with their non-living and living environments". One can hardly disagree, yet they don't seem to go far enough. Seemingly such passive building blocks, one is left wondering if environmental science is more than the sum of its parts, and to what extent it promotes action.



Shark truths and some great white lies

Only seven people this century have been killed by the great white shark, the man-eater of the movies – fewer than the number killed by bee stings. In *Sharks! Predators of the Sea* (Running Press, \$19.98), Piero Angela and Alberto Angela,

aided by 160 full-colour photographs by Alberto Luca Recchi, explore the truths and myths about these ancient predators and reveal why it is more a case of the biter bit, as man's culinary cravings threaten these creatures' survival.

Moving on to another broad and problematic word, pollution, the Longman dictionary describes it as “any harmful or undesirable change in the physical, chemical or biological quality of air, water or soil”, embracing both anthropogenic and natural pollutants. Kemp gives a much longer definition, full of examples. Interestingly, neither introduces aesthetic pollution issues, restricting the idea of unwanted effects to the more physical. As a final awkward concept I looked up Gaia. I was impressed with Kemp’s reasonably unbiased account and his final comment that “the partial or complete removal of mankind might be Gaia’s natural answer to the Earth’s current problems”. The

Longman dictionary gives a tight definition, but no sense of the controversy that surrounds the Gaia hypothesis.

Specialist terms often present problems for the lexicographer, whose expertise can hardly cover every topic. From atmospheric chemistry I took the word immission, which I cannot understand or pronounce, but it was in neither book. Henry’s law is found only in the Longman work, and is given in the dimensioned form with non-SI units ($\text{mol l}^{-1} \text{atm}^{-1}$) that is often used by atmospheric chemists, although others formulate it differently.

The nitrogen oxides are difficult to refer to. In both speech and textbooks they are often introduced carelessly, perhaps because

the word ‘nox’ is so ugly. Longman defines “nitrogen oxides”, “oxides of nitrogen” and “NO_x” (the latter correctly as NO + NO₂). Kemp has no entry for NO_x, which is a pity as a reader might well use this index term. The “oxides of nitrogen”, N₂O, NO and NO₂, have an entry, but are incorrectly equated to NO_x. Oddly, neither dictionary spells sulfur in the way approved by IUPAC.

Overall, however, each of these books will find its readers, and will serve them well, providing either comprehensible short accounts or tight definitions. □

Peter Brimblecombe is in the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

In retrospect chosen by Gordon L. Herries Davies

Principles of Physical Geology

by Arthur Holmes
(1944)

In Britain in the early 1940s things looked pretty black. From Norway and Dunkirk to Crete and Singapore, the war had been a chain of disasters. The public craved the relief of escapism. There was no prospect of victory in sight, but that did nothing to inhibit dreams of a glorious post-war world. The building of castles in the air was easier than the holding of redoubts in Greece or Malaya. The Beveridge Report (1942) offered a vision of British society reborn. The Scott Report (1942) looked forward to the day when there would be national parks for national enjoyment. Butler’s Education Act (1944) held the promise of a new scholastic Elysium.

Up in Durham, Arthur Holmes (1890–1965) was a man of that age. He was the university’s professor of geology and he found himself with time on his hands. By day his teaching duties were diminished because the call to arms had removed most of his students. By night the nocturnal hours dragged on as he performed his air-raid patrols. In the dark skies above, menacing Heinkels opened their bomb doors; in the shadow of St Cuthbert’s shrine below, Holmes could do little more than reflect on the future well-being of his science.

During those troubled years Holmes was by no means unique in his concern for geology’s destiny. The science was widely seen as ailing. For decades it had been plodding through a sludge of stratigraphy within a miasmic fog of fossils. The British geological theatre had lacked real drama ever since 1888 when the curtain fell on that long-running and thrilling production, ‘The Highland Controversy’.

More recent offerings, such as Arthur Vaughan’s epic of 1904 set amid the Carboniferous strata of the Bristol district, might possess great scientific merit, but they were unlikely to bring the audience to its feet. In Victorian times, geology had been a favourite among the sciences, but by the 1930s it was a lame ‘also ran’. Its devotees grumbled about a press which no longer reported on their



Holmes: revitalized interest in geology.

activities. Even the fellowship of the famous Geological Society of London was in decline. Its numbers had fallen from 1,279 in 1922 to 1,067 in 1942, and the society’s president had felt it appropriate to devote half of his 1941 anniversary address to a diagnosis of geology’s malaise. Ten years later I must surely have wondered about the health of the discipline myself, because when I became an undergraduate in 1950 I discovered that three pillars of my first-year reading-list — Rutley’s *Elements of Mineralogy*, Watts’s *Geology for Beginners*, and Woods’s *Palaeontology* — were all written the previous century.

Holmes’s reflections upon the condition of geology brought him to a decision. Through the authorship of a new textbook he would seek to play his part in shaping the subject’s future. That work, *Principles of Physical Geology*, was published during September 1944, its preface bearing the twin dates of July 1942 and May 1944. (By that second date Holmes had moved from Durham to the Regius Chair in Edinburgh.) In retrospect these seem to be singularly apt dates. Individually they mark the eve of those two climacterics of the Second World War, Stalingrad and Normandy. On a very different plane, the appearance of the book stands at an equivalent turning-point in the story

of modern geology. Widely acclaimed, it was soon an international best-seller. Known simply as ‘Holmes’, it was destined to become the vade-mecum of an entire generation of young geologists.

I was given my own first copy by my father sometime during 1945, when I was 13. Displayed in the window of a Manchester bookshop, the volume had excited his interest. That particular copy — a volume from the first of 18 post-1944 reprintings — now lies before me as I write. The claret of the spine has lost its sparkle. The frayed corners, the loose hinges, and the thumb-marked fore-edge all betoken the age and heavy usage of a book which has been my companion through life. It today well merits its honourable retirement from active service, shelved in its Valhalla alongside my Coles, Geikies, Lyells and Ramsays.

Having selected physical geology as his canvas, Holmes created some exciting images. He treated the shaping of the Earth’s topography from arêtes to yardangs. Earthquakes, orogenesis, volcanism, and the cause of ice ages all received attention. Ever since 1911 Holmes had been writing about radioactive minerals and the age of the Earth, so that topic too featured strongly. And the final chapter was decidedly risqué, exploring the notion that our continents might be mobile, an idea which most geologists of the 1940s would surely have dismissed as risible.

Within just a few years of the appearance of that first edition of Holmes, geology began to enjoy a new prosperity, as startling discoveries once again placed the science squarely in the public gaze. The study of an early Holmes must have been a seminal experience for many who subsequently became active participants in the subject’s mid-twentieth-century renaissance. Now, 50 years later, those same geologists are passing from the intellectual scene, and it is fitting that we should remember the text that was for so many their earliest inspiration in the Earth sciences. □

Gordon L. Herries Davies is at Trinity College Dublin and Ballinacloy House, Nenagh, Co. Tipperary, Ireland.

Asynchrony of Antarctic and Greenland climate change during the last glacial period

T. Blunier*, J. Chappellaz†, J. Schwander*, A. Dällenbach*, B. Stauffer*, T. F. Stocker*, D. Raynaud†, J. Jouzel‡, H. B. Clausen§, C. U. Hammer§ & S. J. Johnsen§||

* Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

† CNRS Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), BP 96, 38402 St Martin d'Hères Cedex, Grenoble, France

‡ Laboratoire des Sciences du Climat et de l'Environnement, UMR CEA-CNRS 1572, CEA Saclay, Orme des Merisiers, 91191 Gif sur Yvette, France

§ Department of Geophysics (NBI/APG), University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen O, Denmark

|| Science Institute, Department of Geophysics, University of Iceland, Dunhaga 3, Is-107 Reykjavik, Iceland

A central issue in climate dynamics is to understand how the Northern and Southern hemispheres are coupled during climate events. The strongest of the fast temperature changes observed in Greenland (so-called Dansgaard–Oeschger events) during the last glaciation have an analogue in the temperature record from Antarctica. A comparison of the global atmospheric concentration of methane as recorded in ice cores from Antarctica and Greenland permits a determination of the phase relationship (in leads or lags) of these temperature variations. Greenland warming events around 36 and 45 kyr before present lag their Antarctic counterpart by more than 1 kyr. On average, Antarctic climate change leads that of Greenland by 1–2.5 kyr over the period 47–23 kyr before present.

Greenland ice-core isotopic records^{1,2} have revealed 24 mild periods (interstadials) of 1–3 kyr duration during the last glaciation¹ known as Dansgaard–Oeschger (D–O) events³, where temperature increased by up to 15 °C compared with full glacial values^{4,5}. The temperature over Greenland is essentially controlled by heat released from the surface of the North Atlantic Ocean⁶. This heat is brought to the North Atlantic by the thermohaline circulation which is driven by the North Atlantic Deep Water (NADW) formation⁷. NADW strongly influences the deep circulation on a global scale. Therefore climate change related to the North Atlantic thermohaline circulation is not restricted to the North Atlantic region but has global implications⁸. Data correlating with D–O events have been found, for examples in the southern Atlantic⁹, the Pacific Ocean¹⁰ and probably also the Indian Ocean¹¹. On the continents, related events were found in North America where D–O events are manifested as wetter climate periods^{12,13}. Also, the hydrological cycle in the tropics appears to have changed in parallel with D–O events. This is indicated by changes in atmospheric methane concentration, the main source of which is wetlands situated in the tropics during the last glaciation¹⁴.

Temperature variations in Greenland inferred from the $\delta^{18}\text{O}$ record are characterized by fast warmings and slow coolings during the last glaciation. Antarctic temperature variations (from $\delta^{18}\text{O}$ and δD records) show a different pattern with fewer, less pronounced and rather gradual warming and cooling events over the same period^{15,16}.

One of the great challenges of climate research is therefore to find a mechanism that is able to reconcile the very different dynamical behaviour of high-latitude Northern and Southern hemispheres and their phase relationship during rapid climate change. A crucial test for our understanding of climate change is the relative phase of events in the two hemispheres. Using the global isotopic signal of atmospheric oxygen, Bender *et al.*¹⁷ hypothesized that interstadial events start in the north and spread to the south provided that they last long enough (~2 kyr), postulating that northern temperature variations lead those in Antarctica. On the other hand, they did not exclude the possibility that corresponding events are out of phase, as suggested by Stocker *et al.*¹⁸ on the basis of a simplified ocean–atmosphere model which shows that heat is drawn from the Antarctic region when NADW switches on at the onset of a D–O

event. This leads to a cooling of surface air temperatures in the south and a corresponding warming in the north. Charles *et al.*⁹ used stable-isotope ratios of benthic and planktonic foraminifera in a single core from the Southern Ocean to argue that the Northern Hemisphere climate fluctuations lagged those of the Southern Hemisphere by ~1.5 kyr. They conclude that a direct link of high-latitude climate is not promoted by deep-ocean variability.

Conclusive information regarding the phase lag of climate events can come only from high-resolution palaeoarchives which have either absolute or synchronized timescales with uncertainties smaller than 500 yr (for example, to distinguish the Younger Dryas from the Antarctic Cold Reversal¹⁹, hereafter YD and ACR). This prerequisite is at present satisfied only in polar ice cores. Here we use fast variations in methane during the last glaciation for a synchronization that permits a better estimate of the relative timing of events in the two hemispheres.

The CH_4 records from two Antarctic ice cores (Byrd station 80° S, 120° W; Vostok 78.47° S, 106.80° E) and one Greenland core (GRIP ice core, Summit 72.58° N, 37.64° W) have been used to establish coherent timescales for the Antarctic cores with respect to the GRIP timescale over the period from 50 kyr BP to the Holocene epoch. The timescale for the Greenland reference core (GRIP) has been obtained by stratigraphic layer counting²⁰ down to 14,450 yr BP with an uncertainty of ± 70 yr at 11,550 yr BP (termination of the YD). For older ages, the chronology is based on an ice-flow model²¹. This timescale is not considered to be absolute; for instance, the model-based dating differs from the stratigraphic layer counting in the deeper part of the core by several thousand years. However, our conclusions do not depend on the particular timescale, as explained below.

We are able to show that long-lasting Greenland warming events around 36 and 45 kyr BP (D–O events 8 and 12) lag their Antarctic counterpart by 2–3 kyr (comparing the starting points of corresponding warmings). On average, Antarctic climate change leads that of Greenland by 1–2.5 kyr over the period 47–23 kyr BP. This finding contradicts the hypothesis that Antarctic warmings are responses to events in the Northern Hemisphere¹⁷. The observed time lag also calls into question a coupling between northern and southern polar regions via the atmosphere, and favours a connection via the ocean. Indeed, climate models^{18,22–24} suggest that heat is extracted from the South-

ern Hemisphere when NADW formation switches on. This inter-hemispheric coupling is clearly identified in interstadial events 8 and 12, as well as during the termination of the last glaciation.

Synchronization of ice-core records

Because the porous firn layer on the surface of the ice sheet exchanges air with the overlying atmosphere, air, which gets enclosed in bubbles at 50–100 m below the surface (present-day conditions), has a mean age which is younger than the surrounding ice⁵. Consequently, a timescale for the ice and a timescale for the air need to be known, both of which depend on local climate characteristics. For GRIP, the difference between the age of the ice and the mean age of the gas at close-off depth (Δage) has been calculated using a dynamic model⁵ for firn densification and diffusional mixing of the air in the firn, including the heat transport in the firn and temperature dependence of the close-off density. For Summit, estimates of the main parameters—accumulation rate and temperature^{4,25}—allow the calculation of Δage with an uncertainty of only ± 100 yr between 10 and 20 kyr BP, increasing to ± 300 yr back to 50 kyr BP (ref. 5). Temperature was deduced from the $\delta^{18}\text{O}$ profile as suggested from borehole temperature measurements for the glacial–interglacial transition⁴, corresponding roughly to a linear dependence with a slope of 0.33‰ per °C. This relation is in contrast to today's relation²⁶ with a slope of 0.67‰ per °C. However, there is good evidence that the $\delta^{18}\text{O}$ –temperature relation was close to the one derived from the borehole temperature profile throughout the past 40 kyr (ref. 5).

The model calculations are confirmed by an independent measurement of Δage on the GISP2 core, drilled 30 km to the west of the GRIP drill site. A Δage of 809 ± 20 yr was determined, on the basis of thermal fractionation of nitrogen at the termination of YD²⁷. This is consistent with Δage of 775 ± 100 yr for the YD calculated with the same model used here⁵.

Two steps are necessary to synchronize ice cores via a gas record: (1) correlation of the CH_4 records, and (2) calculation of the age of the ice by determining Δage . The initial GRIP CH_4 profile^{14,28} was refined between 9 and 55 kyr BP by an additional 180 samples measured both in the Bern and Grenoble laboratories, resulting in a mean sampling resolution of about 150 yr in the period of interest. A number of new samples were measured on the Antarctic cores from Byrd and Vostok station; 179 and 56, respectively. These

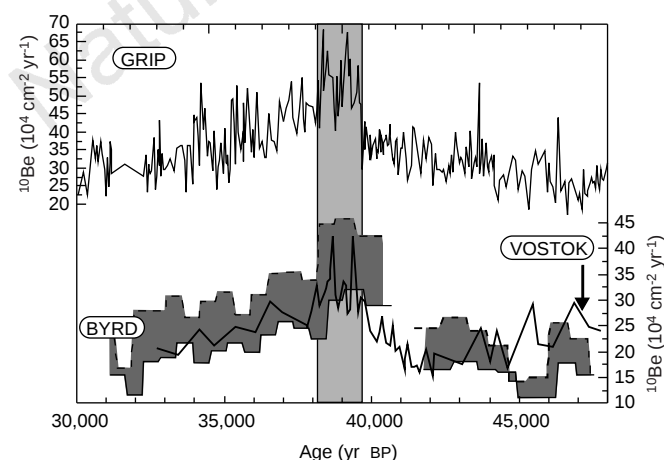


Figure 1 ^{10}Be fluxes from Byrd³², Vostok³³ and GRIP³¹ ice cores. Values for Byrd were reduced by 15%, accounting for a difference in standards used for the measurements on the Byrd and the GRIP/Vostok cores (J. Beer, personal communication). Dating of the GRIP core is after ref. 21. For Byrd, two dating scenarios were used: lower temperature and accumulation (solid line); higher temperature and accumulation (dashed line; see Methods for details). Vostok was dated as described in Methods. The hatched area indicates the ^{10}Be peak area deduced from the GRIP core.

measurements cover the time period 9–47 kyr BP on Byrd and 30–43 kyr BP on Vostok, with a mean sampling resolution of 200 and 250 yr, respectively. For the synchronization of the CH_4 records, a Monte Carlo method was used, searching for a maximal correlation between the records⁵. The fast CH_4 variations allow us to fit the records to ± 200 yr. The time period 25–17 kyr BP was excluded from the synchronization because only low CH_4 variations appear during that period, making the correlation uncertain.

The uncertainty in Δage for the Antarctic sites is dominated by the uncertainty in temperature and accumulation. Estimates of the glacial–interglacial temperature difference at Vostok and Byrd station range from 7 to 15 °C (refs 29, 30). This range is covered in our Δage calculations for Byrd station (see Methods). Δage for the Byrd core is one order of magnitude smaller than for the Vostok core, and comparable to the GRIP Δage . This makes the synchronization of isotopes more convincing for Byrd/GRIP than for Vostok/GRIP (see below and Methods).

To validate the resulting timescales, the distinct double peak of the cosmogenic radioisotope ^{10}Be around 40 kyr BP can be used. The GRIP ^{10}Be record³¹ is plotted together with that of Byrd³² on the new

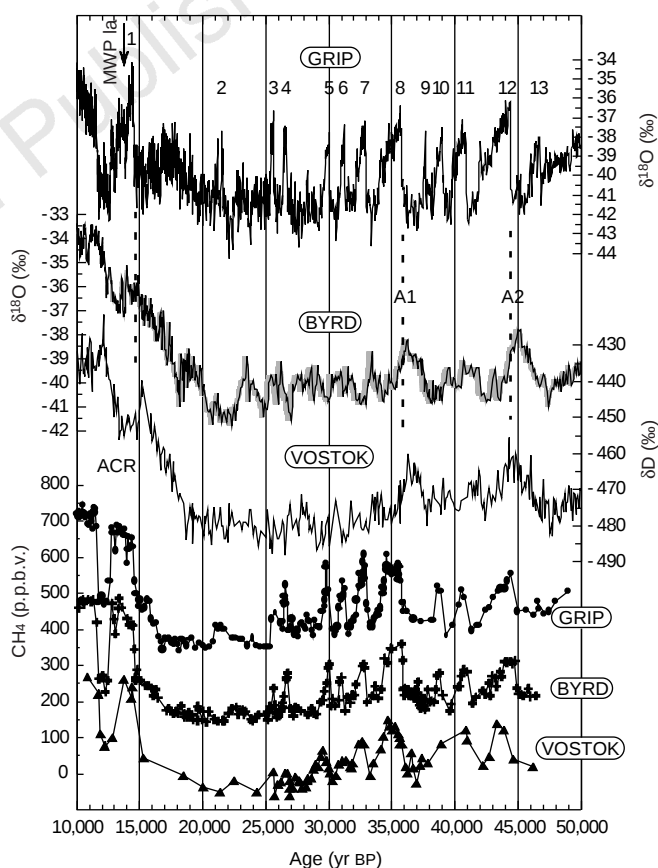


Figure 2 GRIP¹, Byrd¹⁵ and Vostok¹⁶ isotopic and CH_4 records on the common timescale (GRIP timescale in years before 1989). For the data from Byrd, a solid line shows the lower-temperature scenario, and the shaded area indicates the range of Δage for the two temperature scenarios discussed in Methods. The CH_4 scale at lower left corresponds to the GRIP values; Byrd and Vostok values were lowered by 200 and 400 p.p.b.v., respectively, for better visibility. The numbers on top of the GRIP isotopic record indicate the location of Dansgaard-Oeschger events; ACR is the location of the Antarctic Cold Reversal as described by Jouzel *et al.*³⁴; Antarctic warmings are indicated by A1 and A2; vertical dashed lines show the location of Greenland warmings 1, 8 and 12 in the Antarctic cores. The Byrd $\delta^{18}\text{O}$ variation between 25 and 17 kyr BP shows most likely local climate characteristics because: (1) it is not seen in the Vostok or in the Dome C record², and (2) it is not compatible with the climate mechanism seen during events A1, A2 and also the ACR present in all Antarctic records. The arrow marked MWP 1a indicates the position of the first meltwater pulse⁴⁵.

synchronized timescale in Fig. 1. Unfortunately, the resolution of the latter is much lower, which does not allow a firm validation. Nevertheless, the two timescales—resulting from assuming 7 °C or 15 °C glacial–interglacial temperature change—are both consistent with the ^{10}Be signal. As they represent extreme positions relative to the GRIP record around the peak position, we are confident that the true Byrd timescale (relative to GRIP) lies within our estimates. The relative uncertainty at the location of the ^{10}Be peak given by the two temperature models is about 300 yr.

Main features of Antarctic climate

Vostok station has a low accumulation rate and low temperature; the present values are 1.7–2.6 cm of ice per year and –55 °C, respectively¹⁶. This results in a large Δage , and makes Δage extremely sensitive to uncertainties in these parameters. Therefore, the Vostok CH_4 record alone is not well suited to correlate the Northern and the Southern hemispheres on a millennial scale. However, in combination with the detailed ^{10}Be record^{31,33} a reasonable dating relative to Greenland can be achieved (for details, see Methods). The use of Vostok, in view of the relation between Greenland and Antarctic records, is to indicate which features seen in the Byrd core are representative for the climate in Antarctica (we limit our discussion to these features). They include the deglaciation and the ACR, a cold event occurring during the Bølling/Allerød^{19,34,35}, and warm events centring at 35.5 and 44.5 kyr BP (denoted as A1 and A2 in Fig. 2).

Phase lag

In Fig. 2 the isotopic records from Byrd ($\delta^{18}\text{O}$), Vostok (δD) and GRIP ($\delta^{18}\text{O}$), which is representative of Greenland^{2,21} and clearly registers Northern Hemispheric climate change³⁶, are plotted on the

Table 1 Corresponding depths for the three ice cores

Age of the ice (kyr BP)	Depth GRIP (m)	Depth Byrd*† (m)	Depth Vostok† (m)
12	1638.0	1074.4	261.1
13	1674.0	1124.2	278.6
14	1722.6	1161.4	293.1
15	1770.0	1198.6	307.6
16	1802.2	1235.8	322.8
17	1834.3	1269.5	336.4
18	1867.6	1294.4	348.9
19	1894.5	1319.4	360.1
20	1916.5	1344.3	371.0
21	1937.7	1369.3	381.9
22	1958.8	1394.2	392.9
23	1975.6	1419.1	404.3
24	1993.1	1444.1	415.9
25	2011.1	1469.0	427.3
26	2031.6	1493.9	438.9
27	2051.6	1518.7	451.3
28	2065.1	1539.9	464.2
29	2079.9	1561.1	477.6
30	2097.5	1582.3	492.0
31	2112.6	1603.5	505.6
32	2129.5	1624.7	516.4
33	2149.9	1645.8	528.5
34	2162.9	1667.0	541.5
35	2182.7	1688.2	553.8
36	2203.6	1709.4	566.4
37	2214.7	1730.6	581.1
38	2227.5	1751.7	593.5
39	2242.8	1772.9	604.2
40	2255.1	1786.5	614.2
41	2272.0	1798.8	623.6
42	2282.6	1811.2	633.4
43	2297.6	1823.5	642.7
44	2315.9	1840.7	653.2
45	2330.2	1858.3	665.8
46	2340.2	1875.8	680.0
47	2353.2	1893.4	694.2

* The indicated depth corresponds to the higher Byrd temperature scenario (see Methods).

† We note that in the new Byrd, as well as in the new Vostok, age–depth relations, rather unrealistic changes in the annual layer thickness appear. They result because the synchronization is based on the GRIP timescale which is, although well dated, not an absolute timescale. Therefore care must be taken when using the Byrd and Vostok timescales with respect to annual layer or accumulation changes. Our conclusions do not depend on an absolute timescale, and are unaffected by such problems (see Methods).

common timescale (Table 1) based on GRIP. Our synchronization clearly demonstrates that the Antarctic warmings (A1 and A2) are out of phase with the corresponding D–O events (8 and 12). Antarctic temperature variations A1 and A2 are characterized by a gradual rather than abrupt increase and decrease; D–O events by a fast increase and a slow decrease. While the Antarctic temperature is already increasing, Greenland is still cooling. The Antarctic temperature rise is interrupted once Greenland temperature jumps to an interstadial state within only a few decades^{2,21}. Subsequently, the temperature decreases in both hemispheres to full glacial level, where Antarctica reaches this level earlier than Greenland.

We point out that our synchronization excludes the possibility that Antarctic and Greenland events are in phase. To bring the isotopic records in phase, the Byrd Δage would have to be reduced by more than 1,000 yr. As Δage is only about 500 yr during A1 and A2 (see Fig. 3) this would result in a negative Δage , which is impossible. Further, the relative position of A1 is also confirmed by the ^{10}Be peak recorded in the ice of the three investigated cores.

Antarctic isotopic events other than A1 and A2 can hardly be allocated individually with D–O events. However, one can calculate the mean time shift required to obtain the best correlation between Greenland and Antarctic isotopic records in the period 45–23 kyr BP. Best correlations are obtained when the Antarctic signal is shifted towards younger ages by 1–2.5 kyr. This, and the timing of the events (A1 and A2), confirms the chronology obtained by deep sea cores⁹ and contradicts the hypothesis that long interstadials begin in the Northern Hemisphere and spread to Antarctica, creating a lag of Antarctic climate with respect to Greenland (at least for the past 50 kyr).

Implication for north–south connection

A connection between the hemispheres can take place over the ocean or over the atmosphere. Charles *et al.*⁹ suggested that both high northern and southern temperature could be driven over the atmosphere or the ocean by tropical temperature. Whereas coupling between Antarctic and tropical temperatures would be immediate, northern temperature would lag a tropical temperature change due to the thermal influence of the ice sheets or by the ice sheets' influence on salt balance of the North Atlantic surface layer. An immediate response of the Southern Hemisphere to tropical temperature would tend to synchronize the CH_4 (mainly related to tropical moisture changes) and the Antarctic isotope signal, whereas we observe a synchronism of CH_4 and Greenland $\delta^{18}\text{O}$ (refs 5, 14). This makes a coupling via atmosphere improbable, and points instead to a dominant role of the ocean controlling the past climate of both polar regions.

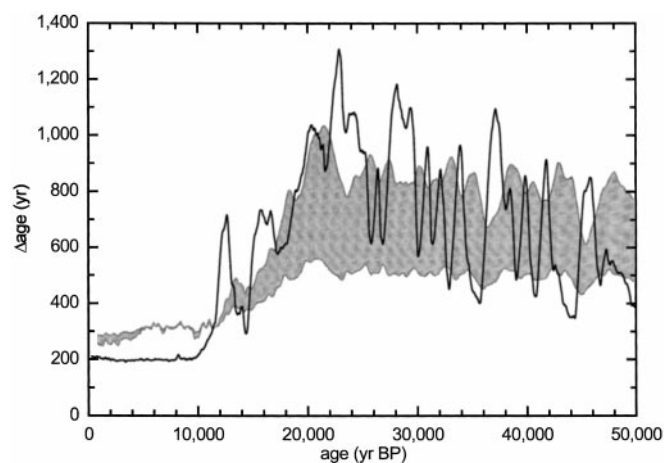


Figure 3 Plot of Δage versus ice age. The solid line shows data from GRIP⁵, the shaded area gives the range of Δage for the Byrd core covered by the temperature scenarios discussed in Methods, calculated with the same model as the GRIP Δage .

It is supposed that changes in the Atlantic thermohaline circulation, the direct cause of temperature changes in central Greenland⁶, are closely linked to ice-rafting events and associated melt water which occur simultaneously in the North Atlantic region³⁷. Melting reduces the deep-water formation and initiates cooling in the North Atlantic region. The cooling then slows the melting and permits re-establishment of the NADW formation, rapidly bringing heat to the North Atlantic⁶. The temperature increase then leads again to a freshwater input into the North Atlantic and to gradual shut down of NADW formation³⁸. Model simulations demonstrate the NADW's sensitivity to freshwater input in the North Atlantic^{24,39}, and they also indicate that the resumption of the 'conveyor belt' circulation after a shut-down is a rapid process^{39,40}.

D–O events and corresponding ice raftings occur more frequently than Antarctic warmings. Therefore, ice-rafting events can only partly, possibly through an increase in sea level, be initiated by Antarctic warmings. For the D–O events that have no counterpart in Antarctica, the trigger mechanism for ice-rafting events thus originates probably from an internal process of one of the ice sheets ablating into the North Atlantic. The Laurentide ice sheet can be excluded as it has a time constant of about 7 kyr (refs 41, 42), longer than the required 2–3 kyr between ice-rafting events³⁷. Bond and Lotti³⁷ proposed instead ice sheets in Greenland, the Barents Sea or Scandinavia as candidates for such a trigger mechanism.

Central to the dynamics that we reconstruct on the basis of the ice cores is the role of the thermohaline circulation in the North Atlantic. Activation of the thermohaline circulation in the North Atlantic tends to cool the Southern Hemisphere^{22,43}. This is exactly what happens during D–O events 12 and 8 (corresponding to A1 and A2), but also during D–O event 1, which is followed by the ACR¹⁹. Further, the YD/ACR out-of-phase relationship between Northern and Southern hemispheres is consistent with model simulations from ref. 18, recently reproduced with a three-dimensional coupled atmosphere–ocean general circulation model²⁴.

We may speculate that Antarctica is asynchronously coupled to the Northern Hemisphere during the whole glacial period investigated. Indeed, some of the smaller D–O events (for example, events 7 or 11) might have a concomitant cooling in Antarctica. The smaller amplitude of these coolings compared to the main events A1 and A2 might point to a positive correlation between the duration of a D–O event and the amplitude of the successive Antarctic cooling. However, it is also possible that the small Antarctic events may merely be noise in the isotopic records. In that case, the heat would only be drawn from Antarctica when Antarctica is in a warm phase. This leads to the speculation that under full glacial conditions Greenland and Antarctic temperature, and hence also thermohaline circulation and Antarctic circumpolar sea surface temperature⁹, are only weakly coupled. A weaker coupling of the hemispheres is achieved when the Atlantic thermohaline circulation exchanges less water with the Southern Ocean, and so has a reduced influence on the heat balance of the southern latitudes. The strength of this circulation also determines its stability to freshwater flux perturbations: a weaker thermohaline circulation is more likely to shut down for a given perturbation⁴⁴. If a weak Atlantic thermohaline circulation switches off, then the north will still experience strong cooling, but with a reduced influence on southern climate. We hypothesize that this is the case during the short D–O events which have no corresponding signal in Antarctic records. D–O events 12 and 8 and the Bølling/Allerød/YD sequence, on the other hand, would be examples of strong coupling between the hemispheres.

Warming in Antarctica tends to reduce sea surface density and hence the density of newly formed deep water. This favours NADW formation⁴³, and hence brings the ocean circulation closer to its strong mode with a tighter coupling of the circumpolar sea surface temperature to the Atlantic thermohaline circulation. A switch-on of the thermohaline circulation after ice-rafting events leads to the observed cooling in Antarctica, which results again in a weaker coupling of the

hemispheres for several thousand years. This is also consistent with an increased heat transport to high northern latitudes during A1 and A2 leading to the observed prolonged D–O events 8 and 12.

During the deglaciation, the climate system shows very similar behaviour. During the first part of the deglaciation the Antarctic warming is not interrupted as no ice-rafting events in the North Atlantic region are taking place: perhaps the lowered sea level makes it impossible for an ice sheet to initiate significant surging of other North Atlantic ice sheets. Then a sudden temperature increase in Greenland (14.5 kyr BP) initiates the same pattern as in the glaciation, namely a cooling in Antarctica (ACR). The first meltwater pulse (MWP 1A)⁴⁵ reduces NADW, at first gradually and then completely, allowing further warming in Antarctica. This makes the deglaciation a complicated process involving Milankovitch forcing, internal ice-sheet processes, and ocean circulation changes.

Further data from Antarctica which are synchronized with the existing records, together with model simulations, will help us disentangle the complicated pattern of climate change between Northern and Southern hemispheres for the entire glacial period. More, and better, data should be available from current European, Japanese and American ice-core drilling programmes. □

Methods

To calculate Δage , the accumulation rate and temperature need to be known. The temperature at the time of precipitation can be determined from the $\delta^{18}\text{O}$ signal recorded in the ice. Assuming that today's spatial dependence between $\delta^{18}\text{O}$ and temperature was also valid in the past, the glacial–interglacial temperature increase in Antarctica is $\sim 7\text{--}10^\circ\text{C}$ (refs 16, 29). On the other hand, recent borehole temperature measurements at Vostok suggest³⁰ a temperature change of up to 15°C .

The amount of snow falling over the year depends on the mean temperature of the inversion layer. This temperature (T_1) can be calculated for the surface temperature (T_G) according to⁴⁶:

$$T_1 = 0.67T_G - 1.2^\circ\text{C} \quad (1)$$

where T_1 and T_G are in $^\circ\text{C}$. Assuming a linear relation between accumulation rate and the derivative of the water vapour partial pressure (P) with respect to temperature (T) (ref. 16), the accumulation rate which has led to a layer now found at depth z is given by:

$$a(z) = a_0 \left(\frac{\partial p}{\partial T} \right)_{T_0}^{-1} \frac{\partial p}{\partial T} \Big|_{T_1(z)} \quad (2)$$

where T_0 and $T_1(z)$ are the inversion layer temperature for today's accumulation rate a_0 and for the past accumulation rate $a(z)$, respectively.

Byrd. As the accumulation is calculated from the temperature and present accumulation rate, its uncertainty is essentially linked to the uncertainty in both parameters. For Byrd we have calculated a new age of the ice for two temperature scenarios. For the warmer scenario, the $\delta^{18}\text{O}$ –temperature relation from ref. 29 was used (equation (3)); for the colder scenario³⁰ we use the linear relation given in equation (4):

$$T = (1.01^\circ\text{C per } \text{‰})\delta^{18}\text{O} + 6.57^\circ\text{C} \quad (3)$$

$$T = (2.07^\circ\text{C per } \text{‰})\delta^{18}\text{O} + 42.7^\circ\text{C} \quad (4)$$

From the two temperature scenarios, accumulation was calculated according to equations (1) and (2) with $a_0 = 11.2\text{ cm per yr ice equivalent}$ ⁴⁷. Thereafter, Δage was calculated using the dynamical model (Fig. 3) leading, in combination with the CH_4 synchronization, to a timescale for the warmer- and the colder-temperature scenario.

The accumulation rate is calculated from these temperatures and therefore also covers a wide range of possible accumulations (mean accumulation rates for the glacial are 7.9 and 5.5 cm yr^{-1} for the higher- and the lower-temperature scenario, respectively). Alternatively, the accumulation rate can be calculated from the annual layer thickness, if the ice flow pattern is known. The flow pattern in the Byrd station area is very complicated (due to bedrock topography and melting at the bottom), which results in uncertainties in the calculated accumulation rate^{15,48}. However, our accumulation range over the glacial is consistent with measured annual layer thickness⁴⁷, assuming a linear thinning of annual layers with depth deduced from the Holocene period.

Vostok. As Vostok is a low accumulation site, the ice from the time period considered here is found in the upper 740 m of the ice sheet. There, ice flow is well known and was described by a glaciological model⁴⁹. Thus, it should in principle be possible to get coherency between the timescales for the ice and the occluded gas, accumulation and temperature. In a first approach the ¹⁰Be peak around 40 kyr was synchronized between GRIP³¹ and Vostok³³. For this purpose, accumulation rates were calculated from equations (1) and (2) from the isotopic temperature¹⁶. Δ age was obtained using this accumulation and the isotopic temperature where the present-day accumulation rate a_0 was set to 2.1 cm yr⁻¹, within the uncertainty for today's accumulation rate⁵⁰. To obtain a timescale from the accumulation rate, annual layer thickness was calculated using the thinning function from the glaciological model⁴⁹. With this approach the main features in the isotopic record, as well as the Vostok CH₄ signal, appear shifted relative to the Byrd record. Reducing the Vostok temperature (suggested by Salamin *et al.*³⁰) and therefore also accumulation rate would make the match even worse. In order to get the Vostok CH₄ record in agreement with its GRIP and Byrd counterpart, Vostok and/or the GRIP/Byrd timescales must be squeezed and stretched.

In a first case, we consider Vostok as reference timescale and adapt the GRIP/Byrd timescales. Consequently accumulation has to be changed in both cores and with the same amplitude and sign; therefore this does not change the relative phasing of GRIP and Byrd. This scenario brings event A2 in to agreement between Byrd and Vostok, but it creates fluctuations in the annual layer thickness of GRIP, which are not expected from the exponential relation between accumulation and $\delta^{18}\text{O}$ (ref. 25).

In a second case, we consider GRIP as the reference timescale. The Vostok gas timescale is squeezed and stretched in order to fit the GRIP CH₄ record. The Vostok ice timescale is treated in the same way which is changing the accumulation rate. Thus Δ age is re-calculated with the accumulation rates consistent with the modified timescale (using the thinning function from the glaciological model⁴⁹). The resulting change in the gas chronology was found to be negligible; this means that the Vostok CH₄ record is still in agreement with the GRIP CH₄ chronology. This scenario also brings event A2 into agreement between Byrd and Vostok, but now creates unexpected fluctuations in the annual layer thickness of Vostok. $\delta^{18}\text{O}$, ¹⁰Be and CH₄ data are presented using this chronology in Figs 1 and 2.

In both cases, maintaining consistency between ¹⁰Be peaks and CH₄ records, we obtain good correlation between the isotope records of Byrd and Vostok, and are therefore confident that the characteristic features reflect identical climate events. We are left with the unsolved problem of annual layer fluctuation either in the Vostok or in the GRIP ice core. However, the synchronization of the Byrd and GRIP core is robust to this problem.

Three cores from Vostok are involved in this study. The isotopic profile was obtained on the 3G core¹⁶ while ¹⁰Be was measured on the 4G core³³. The occurrence of a dust horizon in both cores at around 550 m depth (J. R. Petit, personal communication) allows an exact synchronization of the cores (4G depth increased by 3.41 m). The CH₄ profile was measured on the 5G core, which is of better quality than the 3G and 4G cores. The dust horizon at 550 m depth was not found in the 5G core. We thus performed additional CH₄ measurements on the 3G core over the CH₄ increase associated with D-O event 5, allowing us to estimate the depth offset between 3G and 5G (5G depth decreased by 3 m).

Received 1 September 1997; accepted 13 February 1998.

- Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
- Groote, P. M., Stuiver, M., White, J. W. C., Johnsen, S. & Jouzel, J. Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
- Oeschger, H. *et al.* in *Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takahashi, T.) 299–306 (Vol. 29, Geophys. Monogr. Ser., Am. Geophys. Union, Washington DC, 1984).
- Johnsen, S., Dahl-Jensen, D., Dansgaard, W. & Gundestrup, N. Greenland palaeotemperatures derived from GRIP bore hole temperature and ice core isotope profiles. *Tellus B* **47**, 624–629 (1995).
- Schwander, J. *et al.* Age scale of the air in the summit ice: Implication for glacial–interglacial temperature change. *J. Geophys. Res.* **102**, 19483–19494 (1997).
- Bond, G. *et al.* Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
- Roemmich, D. Estimation of meridional heat flux in the North Atlantic by inverse methods. *J. Phys. Oceanogr.* **10**, 1972–1983 (1981).
- Broecker, W. S. & Denton, G. H. The role of ocean–atmosphere reorganizations in glacial cycles. *Geochim. Cosmochim. Acta* **53**, 2465–2501 (1989).
- Charles, C. D., Lynch-Stieglitz, J., Ninnemann, U. S. & Fairbanks, R. G. Climate connections between the hemisphere revealed by deep sea sediment core/ice core correlations. *Earth Planet. Sci. Lett.* **142**, 19–27 (1996).
- Behl, R. J. & Kennett, J. P. Brief interstadial events in the Santa Barbara basin, NE Pacific, during the

- past 60 kyr. *Nature* **379**, 243–279 (1996).
- Bard, E., Rostek, F. & Sonzogni, C. Interhemispheric synchrony of the last deglaciation inferred from alkenone palaeothermometry. *Nature* **385**, 707–710 (1997).
- Grimm, E. C., Jacobson, G. L. Jr, Watts, W. A., Hansen, B. C. S. & Maasch, K. A. A 50,000-year record of climate oscillations from Florida and its temporal correlation with the Heinrich events. *Science* **261**, 198–200 (1993).
- Benson, L. V. *et al.* Climatic and hydrologic oscillations in the Owens Lake basin and adjacent Sierra Nevada, California. *Science* **274**, 746–749 (1996).
- Chappellaz, J. *et al.* Synchronous changes in atmospheric CH₄ and Greenland climate between 40 and 8 kyr BP. *Nature* **366**, 443–445 (1993).
- Johnsen, S. J., Dansgaard, W., Clausen, H. B. & Langway, C. C. Oxygen isotope profiles through the Antarctic and Greenland ice sheets. *Nature* **235**, 429–434 (1972).
- Jouzel, J. *et al.* Vostok ice core: A continuous isotope temperature record over the last climatic cycle (160,000 years). *Nature* **329**, 403–408 (1987).
- Bender, M. *et al.* Climate correlations between Greenland and Antarctica during the past 100,000 years. *Nature* **372**, 663–666 (1994).
- Stocker, T. F., Wright, D. G. & Mysak, L. A. A zonally averaged, coupled ocean–atmosphere model for paleoclimate studies. *J. Clim.* **5**, 773–797 (1992).
- Blunier, T. *et al.* Timing of the Antarctic Cold Reversal and the atmospheric CO₂ increase with respect to the Younger Dryas event. *Geophys. Res. Lett.* **24**, 2683–2686 (1997).
- Hammer, C. U. *et al.* Report on the Stratigraphic Dating of the GRIP Ice Core (Spec. Rep. of the Geophysical Dept., Niels Bohr Institute for Astronomy, Physics and Geophysics, Univ. Copenhagen, in the press).
- Johnsen, S. J. *et al.* Irregular glacial interstadials recorded in a new Greenland ice core. *Nature* **359**, 311–313 (1992).
- Crowley, T. J. North Atlantic deep water cools the Southern Hemisphere. *Paleoceanography* **7**, 489–497 (1992).
- Stocker, T. F. & Wright, D. G. Rapid changes in ocean circulation and atmospheric radiocarbon. *Paleoceanography* **11**, 773–796 (1996).
- Schiller, A., Mikolajewicz, U. & Voss, R. The stability of the thermohaline circulation in a coupled ocean–atmosphere general circulation model. *Clim. Dyn.* **13**, 325–348 (1997).
- Dahl-Jensen, D., Johnsen, S. J., Hammer, C. U., Clausen, H. B. & Jouzel, J. in *Ice in the Climate System* (ed. Peltier, W. R.) 517–532 (Springer, Berlin, 1993).
- Johnsen, S. J., Dansgaard, W. & White, J. W. C. The origin of Arctic precipitation under present and glacial conditions. *Tellus B* **41**, 452–468 (1989).
- Severinghaus, J. P., Sowers, T., Brook, E. J., Alley, R. B. & Bender, M. L. Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice. *Nature* **391**, 141–146 (1998).
- Blunier, T., Chappellaz, J., Schwander, J., Stauffer, B. & Raynaud, D. Variations in atmospheric methane concentration during the Holocene epoch. *Nature* **374**, 46–49 (1995).
- Robin, G. de Q. in *The Climatic Record in Polar Ice Sheets* (ed. Robin, G. de Q.) 180–184 (Cambridge Univ. Press, London, 1983).
- Salamin, A. N. *et al.* Ice core age dating and paleothermometer calibration based on isotope and temperature profiles from deep boreholes at Vostok Station (East Antarctica). *J. Geophys. Res.* **103**, 8963–8978 (1998).
- Yiou, F. *et al.* Beryllium 10 in the Greenland Ice Core Project ice core at Summit, Greenland. *J. Geophys. Res.* **102**, 26783–26794 (1997).
- Beer, J. *et al.* The Last Deglaciation: Absolute and Radiocarbon Chronologies (eds Bard, E. & Broecker, W. S.) 141–153 (NATO ASI Ser. I 2, Springer, Berlin, 1992).
- Raisbeck, G. M. *et al.* in *The Last Deglaciation: Absolute and Radiocarbon Chronologies* (eds Bard, E. & Broecker, W. S.) 127–140 (NATO ASI Ser. I 2, Springer, Berlin, 1992).
- Jouzel, J. *et al.* The two-step shape and timing of the last deglaciation in Antarctica. *Clim. Dyn.* **11**, 151–161 (1995).
- Sowers, T. & Bender, M. Climate records covering the last deglaciation. *Science* **269**, 210–214 (1995).
- Siegenthaler, U., Eicher, U., Oeschger, H. & Dansgaard, W. Lake sediments as continental $\delta^{18}\text{O}$ records from the transition of glacial–interglacial. *Ann. Glaciol.* **5**, 149–152 (1984).
- Bond, G. C. & Lotti, R. Iceberg discharges into the North Atlantic on millennial time scales during the last deglaciation. *Science* **267**, 1005–1010 (1995).
- Stocker, T. F. & Wright, D. G. The effect of a succession of ocean ventilation changes on radiocarbon. *Radiocarbon* **40**, 359–366 (1998).
- Stocker, T. F. & Wright, D. G. Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature* **351**, 729–732 (1991).
- Wright, D. G. & Stocker, T. F. in *Ice in the Climate System* (ed. Peltier, W. R.) 395–416 (NATO ASI Ser. I 12, Springer, Berlin, 1993).
- MacAyeal, D. R. A low-order model of the Heinrich event cycle. *Paleoceanography* **8**, 767–773 (1993).
- MacAyeal, D. R. Binge/purge oscillations of the Laurentide ice sheet as a cause of the North Atlantic's Heinrich events. *Paleoceanography* **8**, 775–784 (1993).
- Stocker, T. F., Wright, D. G. & Broecker, W. S. The influence of high-latitude surface forcing on the global thermohaline circulation. *Paleoceanography* **7**, 529–541 (1992).
- Tziperman, E. Inherently unstable climate behaviour due to weak thermohaline ocean circulation. *Nature* **386**, 592–595 (1997).
- Bard, E. *et al.* Deglacial sea-level record from Tahiti corals and the timing of global meltwater discharge. *Nature* **382**, 241–244 (1996).
- Jouzel, J. & Merlivat, L. Deuterium and oxygen 18 in precipitation: modeling of the isotopic effects during snow formation. *J. Geophys. Res.* **89**, 11749–11757 (1984).
- Hammer, C. U., Clausen, H. B. & Langway, C. C. Jr Electrical conductivity method (ECM) stratigraphic dating of the Byrd Station ice core, Antarctica. *Ann. Glaciol.* **20**, 115–120 (1994).
- Whillans, I. M. Ice flow along the Byrd station strain network, Antarctica. *J. Glaciol.* **24**, 15–28 (1979).
- Ritz, C. Un Modèle Thermo-Mécanique d'évolution Pour le Bassin Glaciaire Antarctique Vostok-glacier Byrd: Sensibilité aux Valeurs des Paramètres Mal Connus. Thesis, Univ. J. Fourier (1992).
- Jouzel, J. *et al.* Extending the Vostok ice-core record of palaeoclimate to the penultimate glacial period. *Nature* **364**, 407–412 (1993).
- Lorius, C., Merlivat, L., Jouzel, J. & Pourchet, M. A 30,000-yr isotope climatic record from Antarctic ice. *Nature* **280**, 644–648 (1979).

Acknowledgements. This work, in the frame of the Greenland Ice Core Project (GRIP), was supported by the University of Bern, the Swiss National Science Foundation, the Federal Department of Energy (BFE), the Schwerpunktprogramm Umwelt (SPPU) of the Swiss National Science Foundation, the EC program "Environment and Climate 1994–1998", the Fondation de France and the Programme National de Dynamique du Climat of CNRS. We thank F. Finet for the CH₄ measurements on Vostok and part of GRIP, C. Rado and J. R. Petit for ice sampling at Vostok station, C. C. Langway for providing us with additional Byrd samples and F. Yiou, G. Raisbeck and J. Beer for the ¹⁰Be data.

Correspondence and requests for materials should be addressed to T.B. (e-mail: blunier@climate.unibe.ch).

Implications of transposition mediated by *V(D)J*-recombination proteins RAG1 and RAG2 for origins of antigen-specific immunity

Alka Agrawal*, Quinn M. Eastman† & David G. Schatz‡

* Department of Pharmacology, † Department of Molecular Biophysics and Biochemistry, and ‡ Howard Hughes Medical Institute, Section of Immunobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

Immunoglobulin and T-cell-receptor genes are assembled from component gene segments in developing lymphocytes by a site-specific recombination reaction, *V(D)J* recombination. The proteins encoded by the recombination-activating genes, *RAG1* and *RAG2*, are essential in this reaction, mediating sequence-specific DNA recognition of well-defined recombination signals and DNA cleavage next to these signals. Here we show that *RAG1* and *RAG2* together form a transposase capable of excising a piece of DNA containing recombination signals from a donor site and inserting it into a target DNA molecule. The products formed contain a short duplication of target DNA immediately flanking the transposed fragment, a structure like that created by retroviral integration and all known transposition reactions. The results support the theory that *RAG1* and *RAG2* were once components of a transposable element, and that the split nature of immunoglobulin and T-cell-receptor genes derives from germline insertion of this element into an ancestral receptor gene soon after the evolutionary divergence of jawed and jawless vertebrates.

Lymphocytes of the vertebrate adaptive system rely on a diverse array of immunoglobulins and T-cell antigen receptors (TCRs) for specific recognition of antigens. In the germ line, the genes encoding the variable portions of these receptors are typically split into component *V* (variable), *J* (joining) and, in some cases, *D* (diversity) gene segments. One of each type of gene segment is joined together in a site-specific recombination reaction to form the exon that encodes the antigen-binding portion of the polypeptide¹. This reaction, known as *V(D)J* recombination, occurs only in lymphocytes, and in some vertebrate species is responsible for generating much of the diversity seen in antigen receptors. *V*, *D* and *J* gene segments are flanked by recombination signals that consist of conserved heptamer and nonamer sequences separated by a poorly conserved spacer sequence whose length is usually 12 or 23 base pairs (bp) (referred to hereafter as 12-signals and 23-signals, respectively). The signals target the reaction, and efficient recombination only occurs between a 12-signal and a 23-signal, a restriction referred to as the 12/23 rule².

The first phase of the recombination reaction is achieved by the products of the recombination-activating genes, *RAG1* and *RAG2* (refs 3, 4). Together, the *RAG1* and *RAG2* proteins bind two recombination signals, bring them into close juxtaposition (this process is termed synapsis), and cleave the DNA, thereby separating the signals from the flanking coding segments^{5–7}. The DNA-bending high-mobility group proteins HMG1 and HMG2 substantially enhance the efficiency of coordinate cleavage^{8,9}, in part by improving binding to the 23-signal⁹. Cleavage occurs in two steps, with a nick first introduced adjacent to the heptamer to expose a 3'-hydroxyl group on the coding flank, followed by direct nucleophilic attack of the 3'-hydroxyl on the opposite DNA strand^{5,10}. The products are blunt, 5'-phosphorylated signal ends and covalently sealed hairpin coding ends (Fig. 1a). In the second phase of the reaction, the two signal ends are joined precisely to form a signal joint, whereas the coding ends are processed to form a coding joint that typically exhibits both nucleotide loss and addition². An array of ubiquitously expressed

DNA double-strand break repair proteins is essential for the joining phase of the reaction^{11–13}.

There may be several parallels between *V(D)J* recombination and transposition^{14–16}. First, the *RAG1* and *RAG2* genes have a compact genomic organization, as would be expected for components of a transposable element. The two genes lie immediately adjacent to one another and are convergently transcribed in all vertebrate species examined so far, and, in most cases, the open reading frames are not interrupted by introns¹⁷. Second, *V(D)J* DNA cleavage, retroviral integration and transposition proceed through a common pathway that involves exposure of a 3'-hydroxyl group and its attack on a target phosphodiester bond in a magnesium-dependent reaction^{10,18}. Third, the *RAG* proteins remain stably associated with a synapsed pair of recognition elements after DNA cleavage¹⁹, as is common in transposition²⁰. Finally, *RAG1* and *RAG2* can join signal ends to coding ends in a reversal of the cleavage reaction²¹. Here we provide a unifying explanation for these observations and for the split configuration of antigen-receptor genes by showing that the *RAG* proteins catalyse a transposition reaction with striking similarities to those catalysed by other transposases.

An unexpected cleavage product

The cleavage substrates used here were ³²P-body-labelled DNA fragments containing a 12-signal and a 23-signal orientated such that cleavage releases a signal end/signal end (SE/SE) fragment of roughly 320 bp, as well as smaller fragments containing the hairpin coding ends (Fig. 1a). Reactions were done with truncated *RAG1* and *RAG2* proteins substantially purified from the B-lymphoma cell line F2A1 and with purified recombinant HMG2 in a buffer containing magnesium but not energy source. Cleavage products were detectable after 4 min and accumulated for at least 2 h (Fig. 1b), concurring with previous results^{5,6}. In addition, species with a mobility slower than that of the input substrate were visible from 15 min onwards (Fig. 1b). These were not a result of proteins remaining associated with the DNA, because proteins were elimi-

nated with protease digestion and organic extraction before loading of the gel. Two closely spaced bands were visible, with the upper predominating at longer time points. The upper band, hereafter referred to as X, is shown below to be a circular product produced by intramolecular transposition of the SE/SE fragment. The lower band may be a more highly knotted topo-isomer or the result of a one-end strand-transfer event.

Restriction mapping of purified band X DNA indicated that it contained sequences from the SE/SE fragment but not from the left or right arms, and that it was not the result of end-to-end ligation of the SE/SE fragment (data not shown). To begin to study the structure of band X, we incubated cleavage products with *Bal31* nuclease, which is a double-stranded-DNA exonuclease and a single-stranded-DNA exo- and endonuclease. In the presence of the RAG and HMG2 proteins, band X and the SE/SE fragment were substantially more resistant to *Bal31* than were other DNA fragments, whereas all bands were highly nuclease-sensitive after removal of the proteins by protease treatment and organic extraction (Fig. 2a). The protein-dependent nuclease resistance of band X indicated that it, like the SE/SE fragment¹⁹, might be stably associated with the RAG and HMG2 proteins. Indeed, band X DNA could be immunoprecipitated with anti-RAG1, anti-RAG2, and anti-HMG2 antibodies, but not with several control antibodies (Fig. 2b). Thus band X DNA is stably bound by RAG1, RAG2 and HMG2, and this association appears to render the DNA resistant to nuclease digestion.

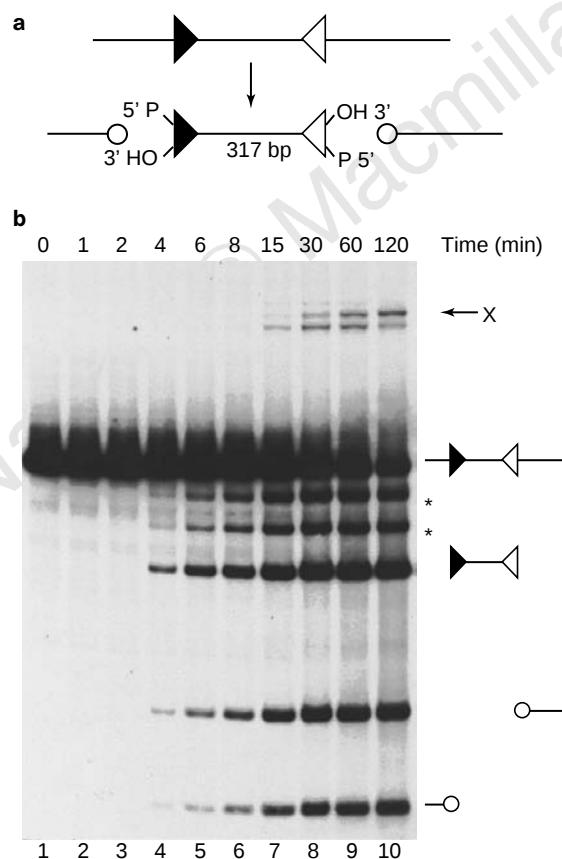


Figure 1 Time course of cleavage and band X formation. **a**, The cleavage substrate was a ³²P-body-labelled DNA fragment, generated by PCR, containing a 12-signal (black triangle) and a 23-signal (white triangle). Cleavage by the RAG proteins generates a 5'-phosphorylated SE/SE fragment and two smaller fragments containing the hairpin coding ends (circles). **b**, Cleavage using RAG and HMG2 proteins was done for different lengths of time at 37 °C and aliquots of the reaction were analysed on a native polyacrylamide gel. Asterisks indicate bands resulting from single cleavage at the 12- or 23-signals.

As band X contained only DNA derived from the SE/SE fragment, we wondered whether band X could be generated using the SE/SE fragment as the starting substrate. Incubation of an SE/SE fragment generated by the polymerase chain reaction (PCR) (Fig. 3a, lane 3) with the RAG and HMG2 proteins generated a low-mobility species

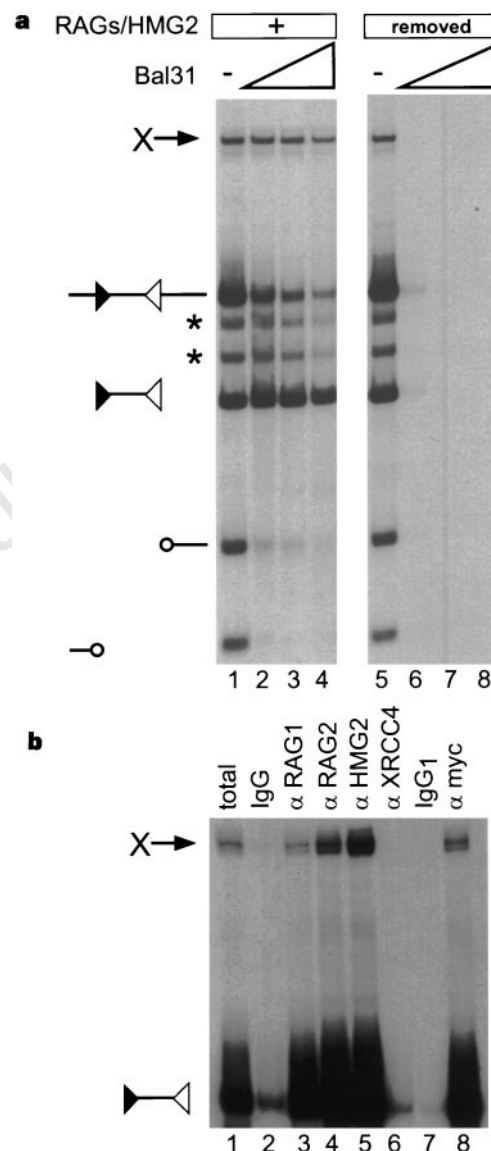


Figure 2 Stable association of the RAG and HMG2 proteins with band X confers nuclease resistance. **a**, Nuclease sensitivity of the cleavage reaction products. After 2 h of cleavage, reaction products were incubated directly with 0, 0.125, 0.5, or 1.0 U *Bal31* nuclease for 1 h at 30 °C in the presence of 2 mM CaCl₂ (lanes 1–4, respectively), or were treated with proteinase K, precipitated and then subject to identical *Bal31* digestions (lanes 5–8). Products were analysed on a native polyacrylamide gel. Asterisks indicate bands resulting from single cleavage at the 12- or 23-signals. **b**, Co-immunoprecipitation of band X DNA with RAG1, RAG2 and HMG2. Band X was generated from the 329-bp SE/SE DNA substrate and immunoprecipitations were done with polyclonal rabbit antibodies (indicated by α) specific for RAG1, RAG2 or HMG2 (lanes 2–4), an immunoglobulin (Ig)G1 monoclonal antibody specific for the Myc epitope tag (lane 8; both RAG1 and RAG2 contain the epitope tag), an IgG1 isotype control monoclonal antibody (lane 7), control polyclonal rabbit IgG (lane 2), or polyclonal rabbit antibodies specific for an irrelevant protein (XRCC4, lane 6). Co-immunoprecipitated DNA was analysed on a native polyacrylamide gel. Lane 1 represents the reaction before immunoprecipitation. As expected¹⁹, the SE/SE DNA substrate is immunoprecipitated by the same antibodies as is band X, with a small amount of nonspecific precipitation seen in lanes 2 and 6.

(Fig. 3a, lane 6) that migrated together with gel-purified band X DNA (Fig. 3a, lane 1) and which was exonuclease-resistant in the presence of the RAG and HMG2 proteins (data not shown). Identical results were obtained using as the substrate an SE/SE fragment purified from a cleavage reaction (Fig. 3b, lane 1). If, however, the substrate lacked one or both of the signals, no product was produced (Fig. 3b, lanes 5, 7, 9). Generation of band X was more rapid when an SE/SE substrate was used rather than a substrate that required cleavage, with products first visible after 8 min of reaction with an SE/SE substrate (data not shown). We conclude that the formation of band X can be dissociated from the process of cleavage and requires a substrate with two signal ends.

The requirement for two signal ends indicated that the RAG proteins, instead of contaminating proteins in the RAG preparation, were probably responsible for formation of band X. To confirm this, we prepared highly purified glutathione *S*-transferase (GST)–RAG

fusion proteins²² and showed that these proteins generate band X (Fig. 3a, lane 8). The weaker activity of this preparation may be explained by its lower specific activity for DNA cleavage (data not shown). These results also showed that formation of band X depended on the addition of HMG2 protein (Fig. 3a, lanes 6, 8), and that HMG2 by itself had no activity (Fig. 3a, lane 9).

Intramolecular transposition

The results above show that a low-mobility DNA species could be generated by RAG1, RAG2 and HMG2 from an SE/SE fragment, but that the reaction does not involve end-to-end ligation to form circles or concatamers. An alternative process that would generate circular DNA molecules without requiring end joining is intramolecular transposition, a strand-transfer reaction in which the free 3'-hydroxyl groups of the signal ends attack phosphodiester bonds in the DNA backbone (Fig. 4a). Such a reaction is a well-characterized feature of bacterial transposases^{23–25} and retroviral integrases^{26,27}, and generates two distinct types of product depending on the topology of the strand-transfer reaction. Attack of the 3'-hydroxyl groups on the opposite strands from which they derive results in an inversion circle in which one portion of the molecule is inverted with respect to the other (Fig. 4a, bottom). In contrast, attack of the 3'-hydroxyl groups on the same strands from which they derive generates two deletion circles, each of which contains one of the recombination signals (Fig. 4a, top). If the attacks occur with a 5' staggered polarity (as is typical for transposition and retroviral integration^{24,28,29}), the two recombination signals should come to be flanked by gaps whose length reflects the staggered positions of strand transfer. When filled in, these gaps should yield target-site duplications, a hallmark of mobile DNA elements³⁰.

On the basis of its discrete mobility in a gel, we predicted that band X would be made up primarily of inversion circles, which should be of uniform size. If this were the case, digestion of band X with a restriction enzyme that cuts once in the SE/SE fragment (for example, *Pst*I) should generate single product with a distinctive structure (Fig. 4a). It should contain all of the sequences from the SE/SE fragment, with one recombination signal and variable amounts of flanking DNA inverted with respect to the rest of the molecule. The two recombination signals should be flanked by gaps derived from the target site. To test these predictions, we cloned *Pst*I-digested band X (Fig. 3a, lane 2) and sequenced the inserts in their entirety. Of the 50 randomly selected inserts analysed, 36 had precisely the structure expected for a transpositionally derived inversion circle and 5 had a structure consistent with a deletion circle (a fragment containing the 12-signal and variable amounts of flanking DNA). The remaining nine clones contained more complex rearrangements or deletions and are not described further here. All 36 inversion-circle clones contained a target-site duplication immediately flanking the recombination signals: 25 (69%) were 5 bp in length, 10 (28%) were 4 bp in length, and 1 (3%) was 3 bp in length (Fig. 4b).

Among the 36 inversion-circle clones, strand transfer occurred at 17 different target sites scattered over the central one-third of the SE/SE fragment (Fig. 4c; target sites for deletion-circle formation are indicated below the line). The paucity of events near the ends may reflect topological constraints or interference from proteins bound to the terminal sequences. The sequences of the target-site duplications and the regions immediately flanking them in the SE/SE fragment are heterogeneous (Fig. 4b), indicating that strand transfer can occur at several different sequences. In addition, the results show that there is one transposition hot spot at which 13 (10 with a 5-bp duplication and 3 with a 4-bp duplication) out of 36 (36%) of the events occurred (Fig. 4c). This indicates that certain target sequences may be preferred over others.

We then studied the structure of the two strands of the inversion circles. The 5' ends of the SE/SE fragment should remain unjoined, resulting in two gaps on one strand. To test this, band X was

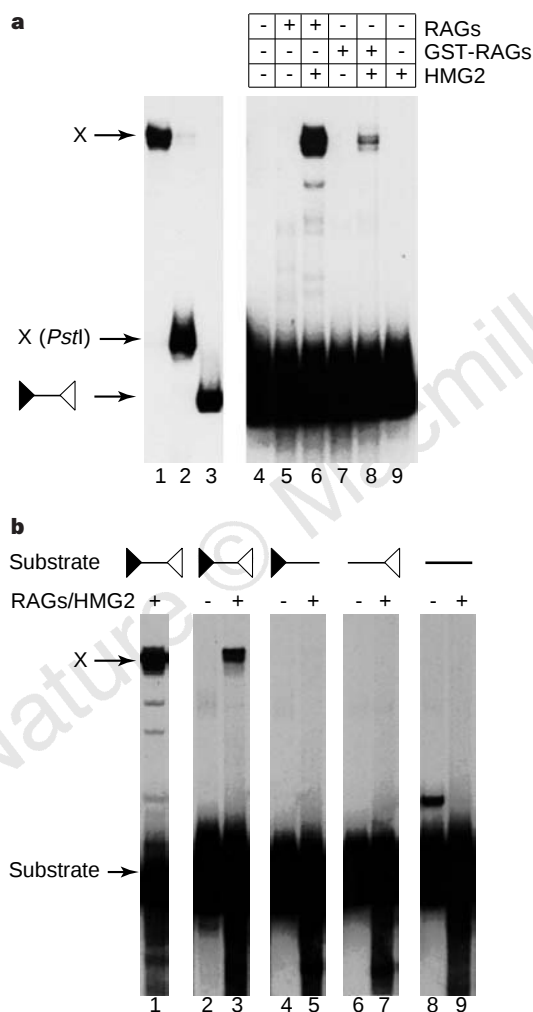


Figure 3 Conditions for band X formation. **a**, A ³²P-end-labelled 329-bp SE/SE DNA substrate was generated by PCR and incubated with combinations of RAG, GST-RAG and HMG2 proteins, as indicated above the lanes, for 2 h at 37°C. Products were analysed on a native polyacrylamide gel. Lanes 1–3 show gel-purified band X, band X digested with *Pst*I, and the SE/SE fragment, respectively. The reduced mobility of *Pst*I-digested band X relative to the SE/SE fragment may be due to the single-stranded gaps in the band X DNA. **b**, ³²P-body-labelled substrates terminating in 12- and 23-signals, a 12-signal only, a 23-signal only, or no signals were generated by PCR (lanes 2–9) or by RAG-mediated cleavage (lane 1) and incubated in the presence or absence of RAG and HMG2 proteins for 2 h at 37°C. The structure of the substrates is indicated above the lanes. Products were analysed on a native polyacrylamide gel.

generated from SE/SE fragments labelled at the 5' end of either the 12-signal or the 23-signal (Fig. 5a) and gel-purified. This material was digested with restriction enzymes that cut close to the ends of the SE/SE fragment (Fig. 5a), and the products were resolved on a denaturing polyacrylamide gel. For each enzyme, band X gave rise to a single band that migrated together with the band produced by digestion of the SE/SE fragment with the same enzyme (Fig. 5b). Therefore, the 5' ends of the SE/SE fragment are unjoined in the band X material, and hence one strand must contain two nicks or gaps. In addition, we confirmed that the other strand was covalently sealed by PCR amplification using divergent primers (depicted as half arrows in Fig. 4a) and gel-purified band X as the template. This reaction yielded a predominant product whose structure was identical to that of inversion-circle clones described above (data not shown). We conclude that the RAG proteins constitute a transposase. Together with HMG2, they can perform intramolecular transposition to yield products that are identical in all major respects to those produced by transposases and retroviral integrases.

Intermolecular transposition

The ability to move to a new genomic location is a defining feature

of transposable elements. To determine whether the RAG proteins could carry out intermolecular transposition, we performed reactions with a 1.4-kilobase (kb) SE/SE substrate containing a tetracycline-resistance gene and a 6.3-kb circular target plasmid containing an ampicillin-resistance gene (Fig. 6a). Transposition should result in a 7.7-kb relaxed circle carrying both antibiotic-resistance genes (Fig. 5a). Direct analysis of the reaction products by Southern blotting showed that substrate sequences were transferred into a product that migrated at the position expected for the relaxed circle (Fig. 6b, lane 5). Formation of this product required the target plasmid and the RAG and HMG2 proteins (Fig. 6b, lanes 2–4). Experiments described below show that at least a portion of this band consists of the expected transposition product; we cannot rule out that it also contains a 'panhandle' product generated by strand transfer of only one end. In addition, substrate sequences also appeared in a band migrating at roughly the position of a linear 7.7-kb product. This band might be generated from relaxed circular molecules by nonspecific nuclease activity, or by concerted transposition of two signal ends derived from different substrate molecules (in which case the product would consist of the linear target plasmid and two copies of the substrate). We ruled out the latter

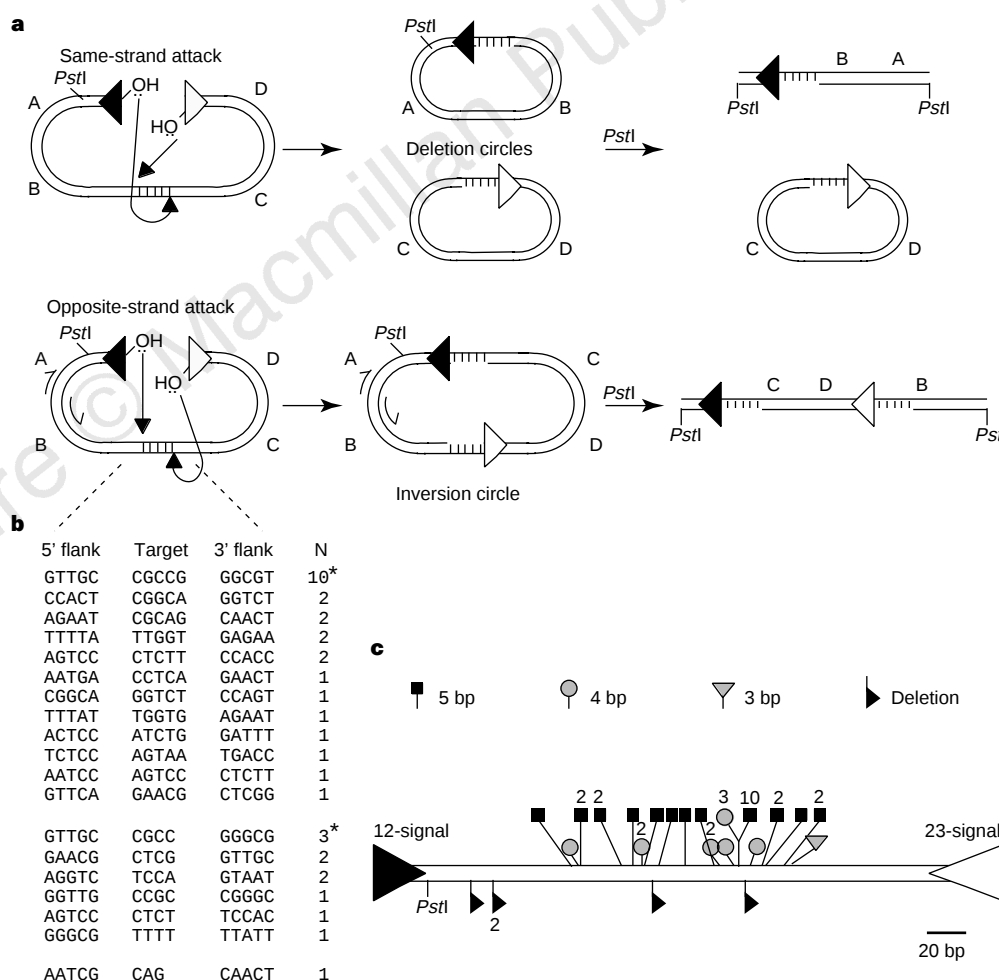


Figure 4 Intramolecular transposition events using the 329-bp SE/SE substrate. **a**, Depending on the topology of the strand attacks, two deletion circles (same-strand attack, top) or a single inversion circle (opposite-strand attack, bottom) are generated during intramolecular transposition. Deletion circles may be catenated and are represented as uncatenated circles for simplicity. Staggered nicks at a 5-bp target site (five short vertical lines) are depicted. This produces 5-bp gaps adjacent to the recombination signals and a 5-bp duplication after repair. The half-arrow indicate the positions of PCR primers used to confirm that one strand of the

inversion circle is intact. **b**, Target-site sequences. *Pst*I-digested band X was cloned and the inserts sequenced. Only inversion-circle target sites and the 5 bp flanking them on either side in the starting substrate are shown. **c**, Location of target sites within the 329-bp SE/SE fragment. Multiple events at a single position are indicated by numbers above or below the symbols. Inversion-circle events are marked above the line (with different symbols used to indicate the length of the target-site duplication) and deletion-circle events below.

explanation by repeating the experiment with smaller target plasmids (<3 kb); here, the linear product contained only one copy of the 1.4-kb substrate (data not shown).

Bacteria were transformed with the reaction products and plated on agar containing ampicillin and tetracycline, or ampicillin alone, and the ratio of the numbers of colonies obtained on these two media provided a measure of the transposition frequency. This indicated that, in the presence of the RAG and HMG2 proteins, 0.016%, or 1 in 6,300, target plasmids were transposition recipients (average of two experiments). A fivefold molar excess of target-to-donor molecules was used in the reaction, indicating that about 1 in 1,300 SE/SE substrates was transposed. The transposition frequency dropped by at least tenfold in the absence of HMG2, and no ampicillin and tetracycline resistant colonies were obtained in the absence of the RAG proteins. Roughly equal numbers of ampicillin-resistant colonies were obtained from the different reactions.

Restriction mapping of plasmids prepared from 38 ampicillin and tetracycline resistant colonies confirmed the presence of the SE/SE

substrate in all of them, and PCR reactions showed that the element was full-length in the ten representative plasmids analysed (data not shown). Sequencing of the two ends of the SE/SE element and flanking target DNA showed that all 38 plasmids contained the expected target-site duplication immediately flanking the recombi-

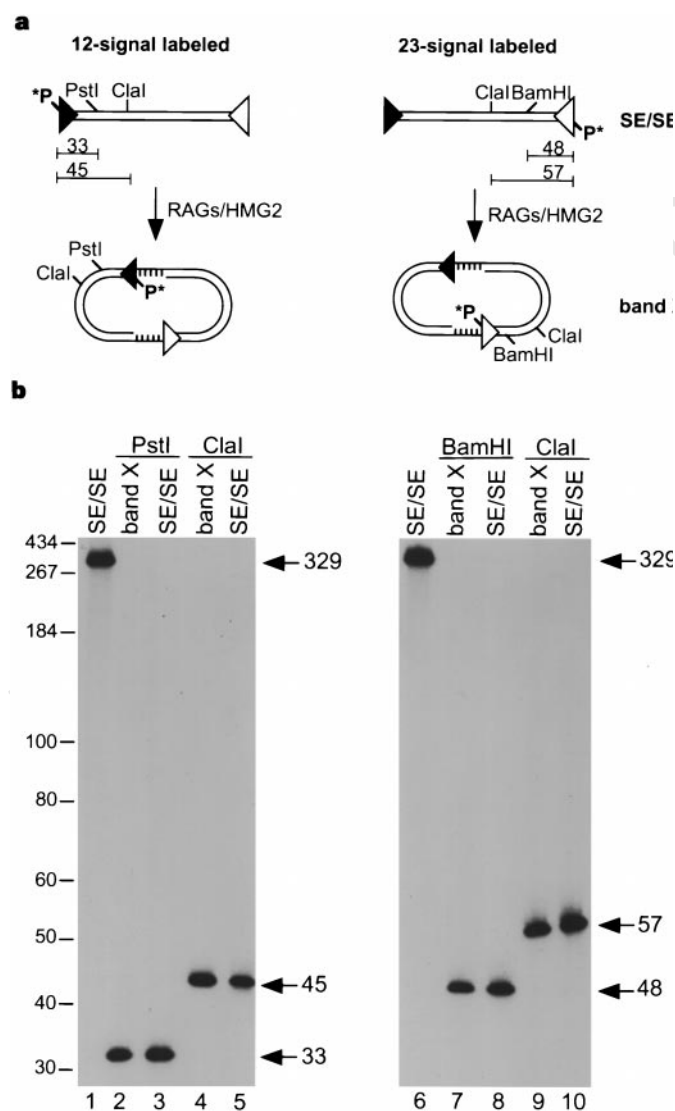


Figure 5 Mapping of the 5' ends of the SE/SE fragment and band X. **a**, A 329-bp SE/SE fragment, 5'-end-labelled with ^{32}P (P*) at either the 12-signal or the 23-signal, was incubated with RAG and HMG2 proteins. The resulting band X products were gel-purified and digested with restriction enzymes that cut close to the labelled ends (distances are indicated in nucleotides). **b**, Digested products were analysed on an 8% denaturing polyacrylamide formamide/urea gel. The sizes of the molecular weight markers are given in nucleotides.

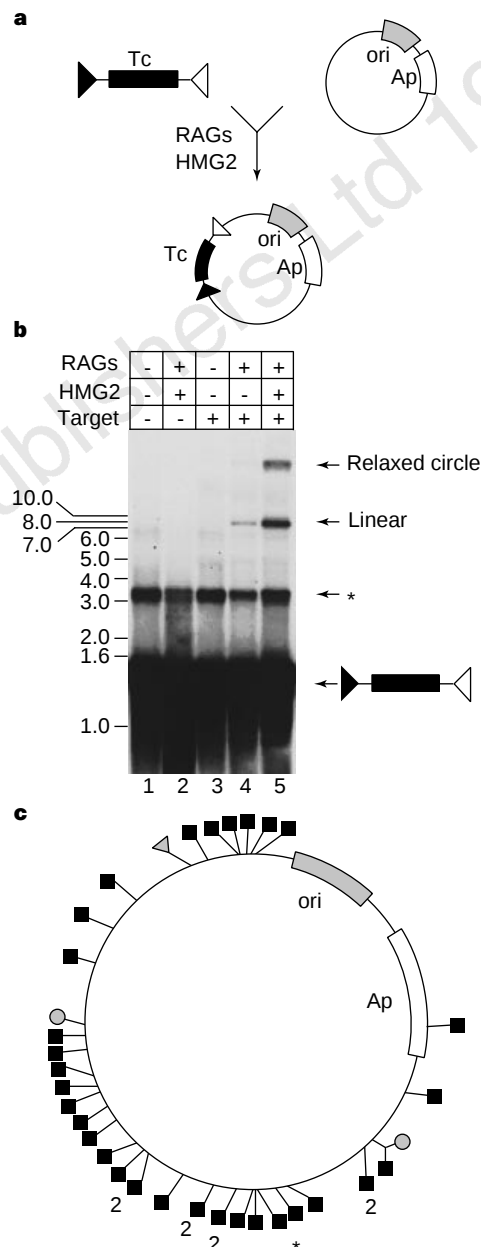


Figure 6 Intermolecular transposition events. **a**, A 1.4-kb substrate containing a tetracycline-resistance gene (Tc) flanked by 12- and 23-signals, and 6.3-kb plasmid target (a mixture of supercoiled and relaxed circular forms) containing an ampicillin-resistance gene (Ap) and a bacterial origin of replication (ori), was incubated with RAG and HMG2 proteins. Intermolecular transposition should generate a relaxed circular plasmid containing both antibiotic-resistance genes. **b**, Southern blot analysis of intermolecular transposition. Reactions were electrophoresed on a 0.7% agarose gel, transferred to a membrane, and hybridized with a ^{32}P -labelled probe generated from the 1.4-kb substrate by random priming. Arrows indicate the positions of the 7.7-kb relaxed circle and linear product. An asterisk indicates the position of a background band present in the substrate preparation. The positions of the molecular weight markers are given in kilobases. **c**, Location of the 38 integration sites in the target plasmid. An asterisk indicates the one event in which target sequences flanking the site of integration had become rearranged. Symbols are as in Fig. 4c.

nation signals. In comparison to the intramolecular reaction, a larger proportion (35 of 38, 92%) of target-site duplications were 5 bp in length, only 2 (5%) were 4 bp in length, and 1 (3%) was 3 bp in length (Fig. 6c). Strand transfer occurred at 33 different sites distributed over much of the target plasmid (Fig. 6c). Target-site sequences were heterogeneous and included GC- and AT-rich regions (data not shown). No hot spot or consensus target sequence was apparent.

We also detected intermolecular transposition using 329-bp or 139-bp SE/SE substrates and a 1.1-kb linear target (data not shown), showing that intermolecular transposition does not require a supercoiled target. In addition, a substrate consisting of an equimolar mixture of double-stranded 12- and 23-signal end oligonucleotides could be transposed into a circular target to generate a linear product (data not shown), indicating that the two signal ends need not reside on the same donor DNA molecule for the reaction to occur. Our results show that the RAG and HMG2 proteins can

mediate intermolecular transposition into diverse target sequences, usually generating a target-site duplication of 5 bp.

Discussion

The transposition reaction mediated by RAG1 and RAG2 is mechanistically similar to that mediated by other transposases. The transposed segment of DNA has terminal inverted repeats (recombination signals), and is excised from a donor site and inserted into a target site on another DNA molecule or within the element itself. Insertion (strand transfer) generates short gaps flanking each end of the transposed element which become target-site duplications when repaired. No energy source is required for the reaction, in keeping with other transpositional strand-transfer reactions³¹.

RAG-protein-mediated transposition may proceed through a series of higher-order protein–DNA complexes analogous to those identified in well-studied bacterial transposition systems^{20,32}.

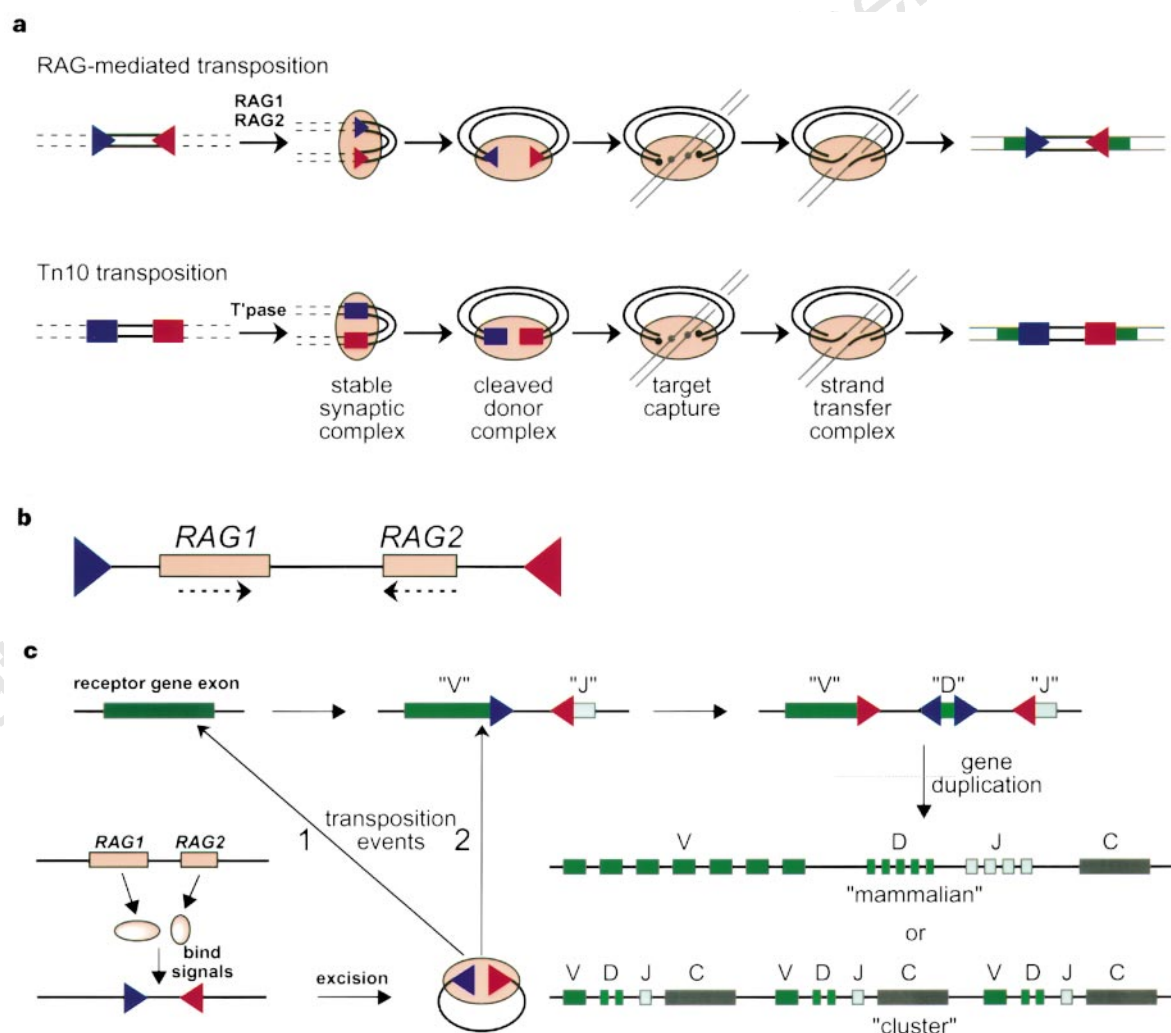


Figure 7 RAG-mediated transposition and a model for the origins of split antigen-receptor genes. **a**, Parallels between RAG- and Tn10-mediated transposition. Tn10 transposase (T'pase) or the RAG proteins (shaded ovals) recognize and synapse the terminal sequences of the transposable element (blue and red triangles or rectangles) and then excise it from donor sequences (dashed lines). Target capture, strand transfer, and gap repair result in target-site duplications (green rectangles) flanking the elements (see text). **b**, Possible structure of the original transposable element that integrated into the germ line of an ancestral vertebrate. Dashed arrows indicate the direction of transcription of the RAG genes. **c**, The present 'split' nature of immunoglobulin and TCR genes is proposed to have arisen from transposition mediated by RAG (bottom left) of one

or two SE/SE elements into a primordial receptor gene exon (top, dark green rectangle), thereby dividing the exon into two or three gene segments, each flanked by one or two recombination signals (blue or red triangles). These gene segments would represent the evolutionary precursors of current V, D and J gene segments (top). Other models for the generation of D segments can also be envisioned. Different patterns of gene duplication (right) would result in the 'mammalian' or 'cluster' configurations of gene segments characteristic of the heavy chain locus of mammals or cartilaginous fishes, respectively. The constant region exons (C) are represented as single grey rectangles. Similar models have been proposed previously^{14,41}.

The first such complex contains two uncleaved recombination signals, the RAG proteins and HMG1 or HMG2 (HMG1/2)³³ ('stable synaptic complex', Fig. 7a). After DNA cleavage, the RAG proteins and HMG1/2 remain tightly associated with the signal ends in a 'cleaved donor complex'^{19,33}. This complex is then capable of target capture and strand transfer to form a 'strand-transfer complex'. The existence of such a complex is suggested by the stable association of the RAG and HMG2 proteins with inversion-circle DNA and by the protein-dependent exonuclease resistance of this DNA species (Fig. 2). Analogous stable strand-transfer complexes have been characterized in bacteriophage Mu and Tn10 transposition^{20,23}. These mechanistic similarities suggest a possible common evolutionary origin for the RAG proteins and other transposases. Although no significant sequence similarities have been found yet to support this idea, we note that the MuA and retroviral integrase catalytic domains adopt similar topologies but have little extended sequence similarity^{34–36}. We emphasize that some aspects of this model for RAG-mediated transposition require experimental confirmation.

Our results extend the range of activities and DNA interactions known to be associated with the RAG proteins. They bind in a sequence-specific way to the recombination signals, and also interact with the flanking coding sequences and with coding ends after cleavage³³. Our results show that RAG proteins bound to a pair of signal ends can capture duplex DNA and transfer the signal end 3'-hydroxyl groups to staggered positions on the two strands of the target DNA. It will be interesting to determine whether interactions with coding flanks, coding ends and transposition target sequences use a single, flexible active site, or whether multiple active sites are involved.

HMG2 is vital for both intramolecular and intermolecular transposition, and may have several distinct functions in the reaction. It probably enhances binding of the RAG proteins to the 23-signal and subsequent formation of a functional 12/23-signal synaptic complex⁹. With short SE/SE substrates, it may stabilize the severe bends in the DNA required for signal-end synapsis and subsequent intramolecular transposition, and, by analogy with the role of integration host factor in Tn10 transposition, may influence target-site selection and the topology of the intramolecular reaction^{32,37}. Retroviral integration occurs preferentially into bent DNA³⁸ and HMG2 may enhance RAG-mediated transposition by modifying the local structure of the target DNA.

Many mobile DNA elements exhibit only limited target-site selectivity³⁸, and, in keeping with this, the target sites of RAG-mediated transposition are heterogeneous in location and sequence. An exception is the hot spot observed in the intramolecular reaction (Fig. 4), which may represent a preferred target-site sequence or may instead be a consequence of the constrained topology of this reaction. Two aspects of target-site selection in RAG-mediated transposition resemble the transposition mediated by Tn10 (refs 32, 39), but not by Tn7 or bacteriophage Mu. First, target-site capture appears to occur after cleavage (Fig. 7a). Second, target-site immunity is not seen (at least under the conditions used), as shown by the robust intramolecular transposition reaction.

Evolution of the vertebrate immune system

All jawed vertebrates studied so far possess adjacent *RAG1* and *RAG2* genes as well as immunoglobulin and TCR genes, which, with a few notable exceptions, must be assembled by somatic recombination before they can be expressed^{16,40}. There is no evidence that any of these molecules, or antigen-specific lymphocytes, are found in jawless vertebrates (hagfish and lamprey) or invertebrates. This indicates that split antigen-receptor genes and the enzymatic machinery necessary for their assembly into functional units arose in the ~100 million years between the divergence of jawless and jawed vertebrates and the divergence of cartilaginous and bony fishes^{16,40}. Our results are evidence in favour of the theory that a vital

event in the evolution of the antigen-specific immune system was the insertion of a 'RAG transposon' into the germ line of a vertebrate ancestor^{14,41}. The RAG transposon would presumably have consisted of recombination signals flanking the *RAG1* and *RAG2* open reading frames (Fig. 7b). Subsequently, the RAG1/RAG2 transposase mobilized the element in the germline and inserted it into an exon of a receptor gene (Fig. 7c). The gene could then only be expressed if the inserted transposon was excised by the RAG proteins and the two ends of the exon rejoined, presumably by ubiquitous DNA double strand break repair factors. This split gene has a structure analogous to that of immunoglobulin light chain and TCR α and γ chain genes. Insertion of a second transposon into the same exon would have resulted in a tripartite structure analogous to that of immunoglobulin heavy chain and TCR β and δ chain genes (Fig. 7c). Subsequent duplication of individual gene segments or of the entire gene would have resulted in the 'mammalian-type' or 'cluster-type' configurations of gene segments observed currently at the heavy-chain locus in mammals and cartilaginous fishes, respectively (Fig. 7c). Continued activity of RAG1 and RAG2 in the germ line could have led to the formation of 'preassembled' genes observed in the immunoglobulin loci of cartilaginous fishes¹⁶.

Our results show that the RAG proteins can perform transposition *in vitro*, and, if the above hypothesis proves correct, then they were once active *in vivo* as well, where they played the determining role in the evolution of the antigen-specific immune system. Do they retain the ability to perform transposition *in vivo*? To our knowledge, no such events have been described, and it is possible that changes in the RAG proteins or some aspect of the cellular environment has led to a suppression of transposition *in vivo*. In this case, it will be interesting to determine how transposition in the cell is regulated. This may reveal some of the evolutionary forces that have resulted in the metamorphosis of a transposase into central components of a site-specific recombinase. □

Methods

Proteins. The RAG1 and RAG2 proteins were partially purified from the cell line F2A1, a mouse B-lymphoma cell line expressing active, truncated forms of the murine RAG1 (amino acids (aa) 264–1,008) and RAG2 (aa 1–387) proteins, both of which contain carboxy-terminal polyhistidine and Myc monoclonal antibody epitope tags. RAG proteins were purified over anion-exchange, Ni²⁺, and heparin columns (Q.M.E., I. Villey and D.G.S., manuscript in preparation). Purity was >60% as judged by silver staining of polyacrylamide gels. GST-RAG1 (aa 330–1,040) and GST-RAG2 (aa 1–388) proteins²² were purified over glutathione-agarose after transient co-transfection of 293-T cells²². Purity was >90% as judged by silver staining of polyacrylamide gels. Murine HMG2 lacking the C-terminal acidic domain and containing an amino-terminal polyhistidine region and protein kinase A phosphorylation site⁴² was expressed in bacteria and purified by thermal denaturation and Ni²⁺ chromatography. Most antibodies used have been described¹⁹. The anti-RAG1 antibody used was P7. The anti-HMG2 antibodies were generated by immunizing rabbits with the peptide CKAGKKPG-RPTGSK, spanning murine HMG2 residues 167–182. Antibodies were affinity-purified against the peptide and exhibit no cross-reactivity towards other proteins, including HMG1 (data not shown). The anti-XRCC4 antibodies were generated by immunizing rabbits with a *Pseudomonas* exotoxin fusion protein containing XRCC4 amino acids 5–263, and were affinity-purified. Immunoprecipitations were done nearly as described¹⁹, except that the buffer used was the transposition reaction buffer plus 0.1% NP40.

DNA substrates. ³²P-body-labelled DNA-cleavage substrates were generated from plasmid pC329 or pC317 by PCR as described^{6,19}. pC317 was generated by inserting the 23-signal of p12 $\times$ 23 (ref. 43) into the *Bam*HI site of p12 (ref. 43). The Tc gene of pBR322 was PCR-amplified using primers that introduced flanking 12- and 23-signals, and cloned to generate pTetRSS. The SE/SE fragment containing the Tc gene was generated from pTetRSS by PCR. All SE/SE fragments were gel-purified before use. The target for intermolecular transposition was pJH104 (ref. 44). Sequences of primers and details of plasmid construction are available on request.

Transposition reactions. Transposition reactions (25 μ l) were performed for 2 h at 37 °C with ~25 ng RAG1 and RAG2 (or 200 ng GST–RAG proteins), 75 ng HMG2 and 0.05 pmol SE/SE fragment or cleavage substrate in a buffer containing 20 mM HEPES–Na⁺, pH 7.5, 10 mM magnesium acetate, 50 mM sodium acetate, 2 mM dithiothreitol, 10 μ M zinc sulphate and 100 μ g ml⁻¹ bovine serum albumin. Reactions were stopped by adding 175 μ l 50 mM Tris–Cl, pH 8, 0.5% SDS, 10 mM EDTA and 20 μ g proteinase K and incubating for at least 1 h at 55 °C. DNA was precipitated, separated on a 4% 29:1 acrylamide:bisacrylamide native polyacrylamide gel in 1 $\times$ TBE buffer, and visualized by autoradiography. Intermolecular reactions were performed for 4 h at 37 °C in the same reaction buffer using 0.05 pmol unlabelled SE/SE fragment and 0.25 pmol target plasmid and included a 15 min preincubation at 37 °C in the absence of target DNA. Reactions were terminated and DNA precipitated as above, and 0.1–1% of the reaction was transformed into electrocompetent MC1061 bacteria and spread on agar plates containing ampicillin (100 μ g ml⁻¹) or ampicillin plus tetracycline (10 μ g ml⁻¹). This yielded an average of 500,000 ampicillin-resistant colonies from 0.1% of the reaction.

Received 6 July; accepted 30 July 1998.

1. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
2. Lewis, S. M. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Adv. Immunol.* **56**, 27–150 (1994).
3. Schatz, D. G., Oettinger, M. A. & Baltimore, D. The V(D)J recombination activating gene (RAG-1). *Cell* **59**, 1035–1048 (1989).
4. Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**, 1517–1523 (1990).
5. McBlane, J. F. *et al.* Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83**, 387–395 (1995).
6. Eastman, Q. M., Leu, T. M. J. & Schatz, D. G. Initiation of V(D)J recombination *in vitro* obeying the 12/23 rule. *Nature* **380**, 85–88 (1996).
7. van Gent, D. C., Ramsden, D. A. & Gellert, M. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* **85**, 107–113 (1996).
8. Sawchuk, D. J. *et al.* V(D)J recombination: modulation of RAG1 and RAG2 cleavage activity on 12/23 substrates by whole cell extract and DNA bending proteins. *J. Exp. Med.* **185**, 2025–2032 (1997).
9. van Gent, D. C., Hiom, K., Paull, T. T. & Gellert, M. Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J.* **16**, 2665–2670 (1997).
10. van Gent, D. C., Mizuuchi, K. & Gellert, M. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271**, 1592–1594 (1996).
11. Weaver, D. T. What to do at an end—DNA double-strand-break repair. *Trends Genet.* **11**, 388–392 (1995).
12. Chu, G. Double strand break repair. *J. Biol. Chem.* **272**, 24097–24100 (1997).
13. Grawunder, U., West, R. B. & Lieber, M. R. Antigen receptor gene rearrangement. *Curr. Opin. Immunol.* **10**, 172–180 (1998).
14. Thompson, C. B. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* **3**, 531–539 (1995).
15. Lewis, S. M. & Wu, G. E. The origins of V(D)J recombination. *Cell* **88**, 159–162 (1997).
16. Litman, G. W. *et al.* Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. *Mol. Biol. Evol.* **10**, 60–72 (1993).
17. Gellert, M. Recent advances in understanding V(D)J recombination. *Adv. Immunol.* **64**, 39–64 (1996).
18. Craig, N. L. Unity in transposition reactions. *Science* **270**, 253–254 (1995).
19. Agrawal, A. & Schatz, D. G. RAG1 and RAG2 form a stable post-cleavage synaptic complex with DNA containing signal ends at V(D)J recombination. *Cell* **89**, 43–53 (1997).
20. Mizuuchi, K. Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.* **61**, 1011–1051 (1992).
21. Melek, M., Gellert, M. & van gent, D. C. Rejoining of DNA by the RAG1 and RAG2 proteins. *Science* **280**, 301–303 (1998).
22. Spanopoulou, E. *et al.* The homeodomain of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. *Cell* **87**, 263–276 (1996).
23. Weinert, T. A., Derbyshire, K. M., Hughson, F. M. & Grindley, N. D. F. Replicative and conservative transpositional recombination of insertion sequences. *Cold Spring Harb. Symp. Quant. Biol.* **49**, 251–260 (1984).
24. Benjamin, H. W. & Kleckner, N. Intramolecular transposition by Tn10. *Cell* **59**, 373–383 (1989).
25. Isberg, R. R. & Syvanen, M. Tn5 transposes independently of cointegrate resolution. Evidence for an alternative model for transposition. *J. Mol. Biol.* **182**, 69–78 (1985).
26. Shoemaker, C., Hoffmann, J., Goff, S. P. & Baltimore, D. Intramolecular integration within Moloney Murine Leukemia Virus DNA. *J. Virol.* **40**, 164–172 (1981).
27. Lee, Y. M. H. & Coffin, J. M. Efficient autointegration of avian retrovirus DNA *in vitro*. *J. Virol.* **64**, 5958–5965 (1990).
28. Fujiwara, T. & Mizuuchi, K. Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**, 497–504 (1988).
29. Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl Acad. Sci. USA* **86**, 2525–2529 (1989).
30. Berg, D. E. & Howe, M. M. *Mobile DNA* (Am. Soc. Microbiol., Washington DC, 1989).
31. Mizuuchi, K. Polynucleotidyl transfer reactions in transpositional DNA recombination. *J. Biol. Chem.* **267**, 21273–21276 (1992).
32. Kleckner, N., Chalmers, R. M., Kwon, D., Sakai, J. & Bolland, S. Tn10 and IS10 transposition and chromosome rearrangements: mechanism and regulation *in vivo* and *in vitro*. *Curr. Top. Microbiol. Immunol.* **204**, 49–82 (1996).
33. Hiom, K. & Gellert, M. Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol. Cell* **1**, 1011–1019 (1998).
34. Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986 (1994).
35. Rice, P. & Mizuuchi, K. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* **82**, 209–220 (1995).
36. Bujacz, G. *et al.* High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* **253**, 333–346 (1995).
37. Chalmers, R., Guhathakurta, A., Benjamin, H. & Kleckner, N. IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. *Cell* **93**, 897–908 (1998).
38. Craig, N. L. Target site selection in transposition. *Annu. Rev. Biochem.* **66**, 437–474 (1997).
39. Sakai, J. & Kleckner, N. The Tn10 synaptic complex can capture a target DNA only after transposon excision. *Cell* **89**, 205–214 (1997).
40. Rast, J. P. *et al.* α , β , γ and δ T cell antigen receptor genes arose early in vertebrate phylogeny. *Immunity* **6**, 1–11 (1997).
41. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280**, 288–294 (1979).
42. Zwilling, S., König, H. & Wirth, T. High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. *EMBO J.* **14**, 1198–1208 (1995).
43. Lewis, S. M. & Hesse, J. E. Cutting and closing without recombination in V(D)J joining. *EMBO J.* **10**, 3631–3639 (1991).
44. Hsieh, C., McCloskey, R. P., Radany, E. & Lieber, M. R. V(D)J recombination: evidence that a replicative mechanism is not required. *Mol. Cell. Biol.* **11**, 3972–3977 (1991).

Acknowledgements. We thank D. Hesslein for GST–RAG protein; I. Viley for HMG2 and RAG proteins; C.-L. Tsai for RAG proteins; P. Shockett for help with denaturing polyacrylamide gels; W. Stephen for bacterial plates and media; F. Livak for pTetRSS; E. Spanopoulou for the GST–RAG plasmids; T. Wirth for the HMG2 plasmid; J. Repasky and E. Corbett for help with various procedures; N. Craig and G. Litman for comments on the manuscript; and the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University for oligonucleotides and DNA sequencing. D.G.S. is an associate investigator of the Howard Hughes Medical Institute. This work was supported by a grant from the NIH.

Correspondence and requests for materials should be addressed to D.G.S. (e-mail: david.schatz@yale.edu).

Tidal disruption of the Magellanic Clouds by the Milky Way

M. E. Putman*, B. K. Gibson*, L. Staveley-Smith†, G. Banks‡, D. G. Barnes†, R. Bhatall, M. J. Disney‡, R. D. Ekers†, K. C. Freeman*, R. F. Haynes†, P. Henning†, H. Jerjen*, V. Kilborn§, B. Koribalski†, P. Knezek#, D. F. Malin*, J. R. Mould*, T. Oosterloo**, R. M. Price†, S. D. Ryder††, E. M. Sadler‡‡, I. Stewart†, F. Stootman||, R. A. Vaile||§§, R. L. Webster§ & A. E. Wright†

* Mount Stromlo & Siding Spring Observatories, Australian National University, Weston Creek PO, Weston, ACT 2611, Australia

† Australia Telescope National Facility, CSIRO, PO Box 76, Epping, New South Wales 2121, Australia

‡ University of Wales, Cardiff, Department of Physics & Astronomy, PO Box 913, Cardiff CF2 3YB, UK

§ University of Melbourne, School of Physics, Parkville, Victoria 3052, Australia

|| University of Western Sydney Macarthur, Department of Physics, PO Box 555, Campbelltown, New South Wales 2560, Australia

¶ University of New Mexico, Department of Physics & Astronomy, 800 Yale Blvd. NE, Albuquerque, New Mexico 87131, USA

The Johns Hopkins University, Department of Physics & Astronomy, 34th & North Charles Streets, Baltimore, Maryland 21218, USA

☆ Anglo-Australian Observatory, PO Box 296, Epping, New South Wales 2121, Australia

** Istituto di Fisica Cosmica, via Bassini 15, I-20133 Milano, Italy

†† Joint Astronomy Center, 660 North Aohoku Place, Hilo, Hawaii 96720, USA

‡‡ University of Sydney, Astrophysics Department, School of Physics, A28, Sydney, New South Wales 2006, Australia

Interactions between galaxies are common, and influence physical properties such as the global morphology and star-formation rate¹ (Hubble type). Galaxies can interact in many different ways: they can merge together; they can pass through each other, with gas being stripped from the smaller of the two and compressed in the larger; and they can interact gravitationally² (including, for example, tides in clusters). The relative importance of these mechanisms is often not clear, as the strength of each depends on poorly known parameters such as the density, extent and nature of the dark-matter haloes that surround galaxies³. A nearby example of a galaxy interaction where the mechanism is controversial is that between our Galaxy and two of its neighbours, the Magellanic Clouds. Here we present the results of an atomic-hydrogen survey that help to elucidate this mechanism. Our data reveal a new stream of gas that lies in the opposite direction to the trailing Magellanic Stream and leads the motion of the Clouds. The existence of both leading and trailing streams supports a gravitational interaction whereby the streams are torn from the bodies of the Magellanic Clouds by tidal forces.

The Magellanic Stream, discovered 25 years ago⁴, is a narrow (~10° wide) tail of neutral hydrogen, which trails the Magellanic Clouds along a circle around the Milky Way. The past two decades have seen the number of viable mechanisms for its formation reduced to two—ram-pressure stripping of material from the Magellanic System during its passage through the Galactic halo or extended ionized disk^{5–8}, and tidal interaction of the Magellanic Clouds with the Milky Way^{9–14}. Although tidal models have been successful in explaining many characteristics of the Stream, they have suffered from two major drawbacks: the absence of stars in the Stream¹⁵ (stars are also affected by tidal forces), and lack of evidence for a leading stream, or arm, which is a natural consequence of tidal

interaction^{9–12}. Discrete neutral hydrogen clouds on the leading side have been known for some time, but have generally been regarded as a separate group of high velocity clouds^{5,6}.

The H I Parkes All-Sky Survey (HIPASS)¹⁶ is a new survey for H I in the southern sky ($\delta \leq 0^\circ$) over the velocity range $-1,200$ to $12,700 \text{ km s}^{-1}$. It therefore spans the Milky Way, the Magellanic Clouds and the more distant Universe. The 64-m Parkes telescope, with a focal-plane array of 13 beams, set in a hexagonal grid, is being used to survey the sky in 8° zones of declination, with approximate Nyquist sampling. The spectrometer has 1,024 channels for each polarization and beam, with a velocity spacing of 13.2 km s^{-1} and a spectral resolution, after Hanning smoothing, of 26.4 km s^{-1} . Although not ideal for resolving fine spectral structure, accurate column densities and velocity fields are obtained. The HIPASS survey will scan the entire southern sky five times for full sensitivity. The data presented here come from the first scan only and cover the sky south of declination -62° . The r.m.s. brightness temperature sensitivity is approximately 20 mK, corresponding to a column density sensitivity of $10^{18} \text{ atoms cm}^{-2}$ in each channel. The data were reduced using a modified version of the standard HIPASS reduction software¹⁷ (which was designed for imaging discrete H I sources). The bandpass correction was calculated for each beam and velocity by breaking each 8° scan into five sections, finding the median emission in each section and using the minimum of the five values. This greatly increases our sensitivity to large-scale structure without significant loss of flux density, except near the Galactic Plane.

Figure 1 is a detailed image of the Magellanic System. This image shows the H I peak intensity distribution in a 2,400 square-degree mosaic, centred on the South Celestial Pole. Components such as the Large Magellanic Cloud (LMC), Small Magellanic Cloud (SMC), Magellanic Bridge, Galactic Plane and beginning of the

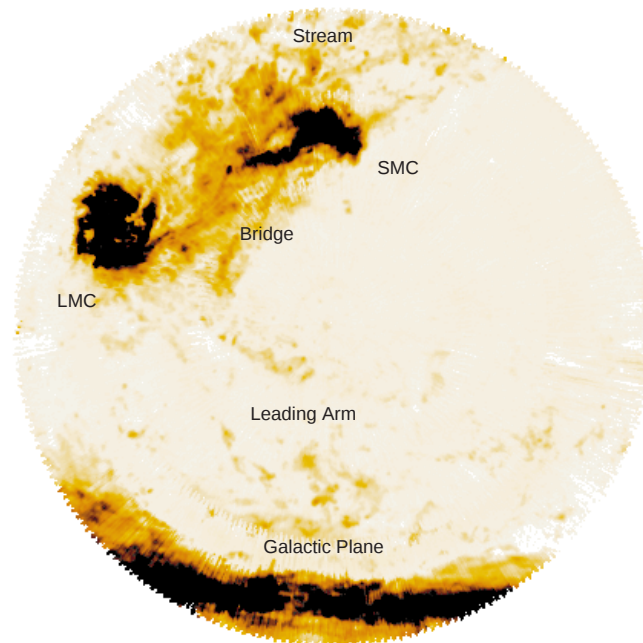


Figure 1 The 2,400 square-degree mosaic of the South Celestial Pole, with the main features labelled, including the area containing the newly identified Leading Arm. The Stream extends 100° beyond its labelled starting point. This image is a peak neutral hydrogen intensity map using the first scans of HIPASS data (see text) in the velocity range $v_{\text{lsr}} \approx 82$ to 400 km s^{-1} . Declination -90° is at the centre, with the edge of the image extending to declination -62° . Right ascension 0 h is at the top of the figure and increases anticlockwise. Note the lack of H I emission in the region between the SMC and Leading Arm. Discontinuity of emission at the Galactic Plane is an artefact of the bandpass correction.

§§ Deceased.

Magellanic Stream are identified. The feature we wish to focus on here however, lies 'between' the Clouds and the Galactic Plane and is labelled 'Leading Arm'. The Leading Arm is shown in more detail in the channel maps of Figs 2 and 3. Of particular interest is the connection of the Leading Arm to the Magellanic Clouds shown in Fig. 2 and, despite its relative thinness ($\sim 1/4$ the width of the trailing Stream), the continuity of the Arm to at least Galactic latitude $b \sim -8^\circ$. The link between the Magellanic System and the strong emission features at $(l, b) = (297^\circ, -24^\circ)$ and $(l, b) = (302^\circ, -16^\circ)$, is not visible in earlier data of Mathewson and Ford¹⁸ (their Fig. 2) or of Morras¹⁹, because of their sparse spatial sampling. Indeed, it is the lack of continuity in the earlier data that was offered as evidence for the absence of a leading arm^{6,8}.

The velocity distribution of the leading H I also suggests a continuous flow of material that originates from the Magellanic System. Figure 2 steps through the velocity channels from $v_{\text{LSR}} = 171$ to 356 km s^{-1} . The H I emission is first seen at the position of the SMC and then continues into the bridge and the LMC. There is subsequently a flow in velocity 'towards' the Galactic Plane through the newly discovered Leading Arm, with the emission disappearing at $\sim 356 \text{ km s}^{-1}$. In this direction $(l, b) = (300^\circ, 0^\circ)$, the maximum Galactic disk velocity allowed under the assumption of circular rotation is $\sim 120 \text{ km s}^{-1}$ (ref. 20), in agreement with the observed upper limit²¹. The observed velocity of the Leading Arm ($\sim 310 \text{ km s}^{-1}$; Figs 2, 3), coupled with its velocity-space continuity with the Magellanic System, clearly eliminates the possibility that the material is associated with the Milky Way disk.

This continuous Leading Arm, extending $\geq 25^\circ$ from the Magellanic System, is a natural prediction of tidal models⁹⁻¹⁴. The most sophisticated of such models to date¹⁰ predicts a ~ 1.5 Gyr-old arm with a mass $\sim 1/3$ that of the Stream, leading the Magellanic Clouds and spanning the region $l \approx 280^\circ$ to 310° and $b \approx -30^\circ$ to $+50^\circ$. When considering the region $b < 0$ (that is, the region shown in Figs 1-3), the mass of the predicted arm is $\sim 1/8$ that of the Stream, with

a relatively flat velocity distribution and a deviation from the Great Circle defined by the trailing Stream of $\sim 30^\circ$. Some care must be taken when directly comparing these model results with new observational data, as three-dimensional spatial information does not exist. Under the assumption of constant and equal distances, the mass ratio of the observed Leading Arm to the Stream is $\sim 1/25$. However, there is reason to believe that the tip of the Stream is closer than its head^{11,12}, implying that the inferred observed mass ratio obtained here is actually a lower limit. The velocity and projected orientation of the observed Leading Arm lie within $\sim 50 \text{ km s}^{-1}$ and $\sim 30^\circ$, respectively, of that predicted by Gardiner and Noguchi¹⁰. This overall agreement is remarkable when the simplifying assumptions employed and the limited parameter space covered are considered (in particular, the shape of the LMC potential and the mass of the Galactic Halo). Our recognition of the missing Leading Arm feature emphasizes the need to advance current tidal models and address the remaining discrepancies.

The identification of a continuous Leading Arm to the Stream, of at least 25° in length, argues against the ram pressure model, but with two caveats. First, in any tidal model, stars are expected to be torn from the Magellanic System along with the gas, yet so far searches for stars associated with the Stream have been negative^{15,22}. This may be due to previous searches not being sensitive to an enhancement of stellar populations older than the ~ 1.5 Gyr Stream age favoured by current tidal models^{10,23}. More probably it is because the initial gas distribution is significantly more extended than the stellar component, as confirmed in interferometer studies of spiral galaxies²⁴ and as seen in other interacting systems²⁵. A second caveat is that some ram pressure models form leading features^{7,8}, but only at the expense of significantly worsening the predicted extent and velocity profile of the Magellanic Stream.

In Fig. 4, we present an expanded view of the LMC/SMC system. This view is remarkable in showing the LMC to have a much more regular spiral structure in neutral hydrogen than it does optically²⁶.

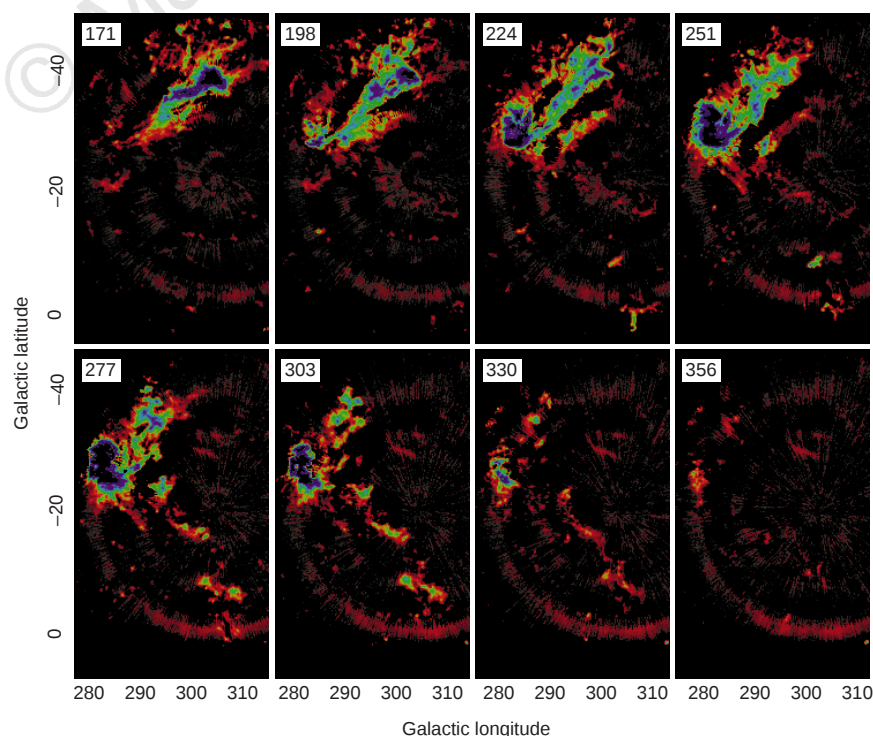


Figure 2 Channel maps of the Magellanic System as labelled in Fig. 1; v_{LSR} is labelled in km s^{-1} in the left upper corner of each channel. Note the continuous flow of the H I emission starting from the SMC and moving into the Bridge, the LMC and finally, the newly discovered Leading Arm. The background striping is a result of the scanning method used for the HIPASS survey. The high positive

velocity of the arm eliminates the possibility of stray radiation effects²⁹ and an association with material from the Milky Way disk^{20,21}. For comparison, the Magellanic Stream extends from 100 km s^{-1} near the Clouds, to $\sim 200 \text{ km s}^{-1}$ at its tip⁴. The orientation of this figure was chosen to match Fig. 1.

There is also evidence for a thin tidal tail emanating from the LMC at $(l, b) = (284^\circ, -33^\circ)$ and extending into the Bridge region towards the SMC. Tentative evidence for such a tail has been presented before²⁷, but Fig. 4 provides unequivocal evidence for such a tail and thus for a mutual tidal interaction between the LMC

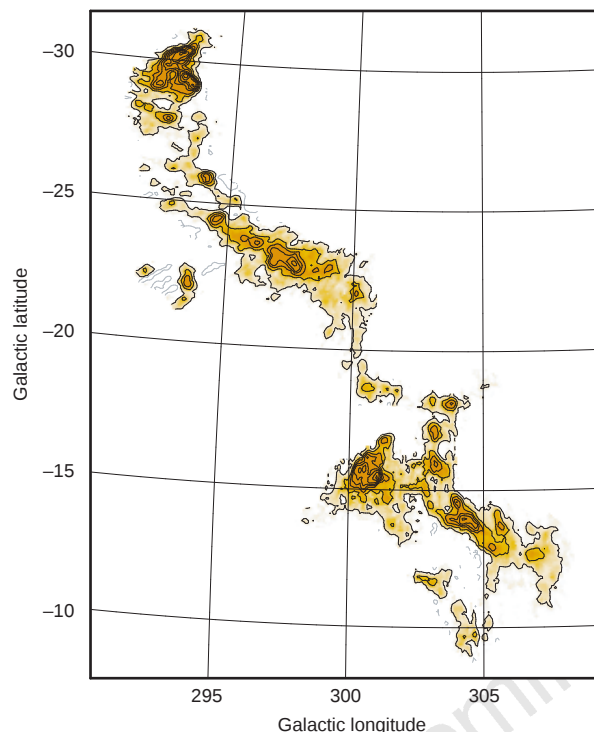


Figure 3 A detailed view of the Leading Arm at the channel centred upon $v_{\text{lsr}} = 323 \text{ km s}^{-1}$. Contours are from 10 to 90% of the brightness temperature maximum ($T_B = 0.88 \text{ K}$). Contrast the continuous nature of the Arm, with the previously assumed 'discrete' nature of the H I in Fig. 2 of ref. 17. The total H I mass of the Arm is $\sim 1 \times 10^7 M_\odot$, and it deviates from the Great Circle defined by the trailing Stream by $\sim 60^\circ$. Both the deviation angle and the H I mass are consistent with model predictions¹⁰. The orientation of this figure was chosen to match Fig. 1.

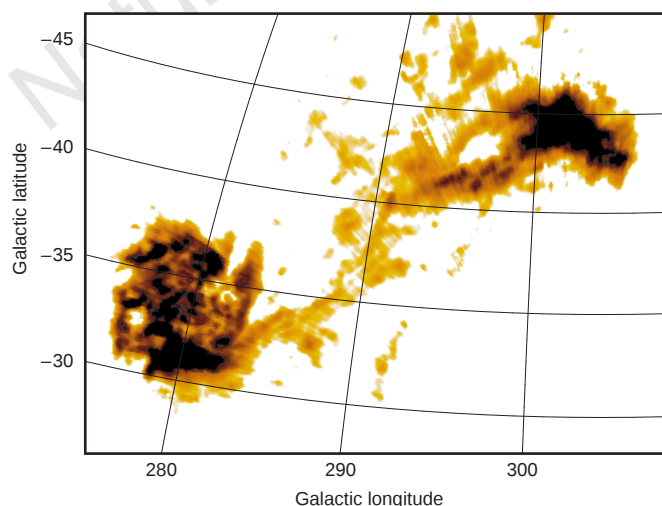


Figure 4 Expanded view of the LMC + SMC + Bridge regions from Fig. 1. The LMC is at the bottom left, with the bridge extending to the position of the SMC at the top right. Note the emission filament emanating directly from the LMC (at $(l, b) = (284^\circ, -33^\circ)$) into the Bridge. This suggests that the Bridge may contain more mass from the LMC than was previously thought¹¹. The orientation of this figure was chosen to match Fig. 1.

and SMC. This is in addition to the evidence presented in Figs 1–3 for a tidal interaction between the Galaxy and the Magellanic System as a whole.

In the future, it will be interesting to determine the full extent of the Leading Arm. Does there exist a complete ring of Magellanic Cloud material? A recent metallicity determination²⁸ for a high-velocity cloud lying above the Galactic Plane at $(l, b) = (287^\circ, 23^\circ)$ suggests a Magellanic Cloud origin, possibly implying a greater extent for the Leading Arm. Exploring the velocity structure of the Leading Arm in more detail will also aid our understanding of the Milky Way's interaction with the Magellanic Clouds. These types of observation provide crucial constraints on simulations aimed at reproducing the formation and evolution of the Magellanic System. This remains one of the only extragalactic systems for which an approximate three-dimensional orbit has been measured and it is an invaluable probe of the mass and extent of the Galactic halo. A full understanding of the role of accretion and tidal interaction in this local 'Rosetta Stone' is essential if we are fully to appreciate their effects in the Universe at large. $\square$

Received 16 March; accepted 9 July 1998.

1. Hashimoto, Y., Oemler, A., Lin, H. & Tucker, D. L. The influence of environment on the star formation rates of galaxies. *Astrophys. J.* **499**, 589–599 (1998).
2. Barnes, J. E. & Hernquist, L. Dynamics of interacting galaxies. *Annu. Rev. Astron. Astrophys.* **30**, 705–742 (1992).
3. Carr, B. Baryonic dark matter. *Annu. Rev. Astron. Astrophys.* **32**, 531–590 (1994).
4. Mathewson, D. S., Cleary, M. N. & Murray, J. D. The Magellanic Stream. *Astrophys. J.* **190**, 291–296 (1974).
5. Mathewson, D. S., Wayte, S. R., Ford, V. L. & Ruan, K. The 'High Velocity Cloud' origin of the Magellanic Stream. *Proc. Astron. Soc. Aus.* **7**, 19–25 (1987).
6. Barnes, J. E. in *Formation of the Galactic Halo* (eds Morrison, H. & Sarajedini, A.) 415–423 (ASP Conf. Ser., San Francisco, 1996).
7. Sofue, Y. Fate of the Magellanic system. *Publ. Astron. Soc. Japan* **46**, 431–440 (1994).
8. Moore, B. & Davis, M. The origin of the Magellanic Stream. *Mon. Not. R. Astr. Soc.* **270**, 209–221 (1994).
9. Mura, T. & Fujimoto, M. The Magellanic Stream and the Galaxy with a massive halo. *Publ. Astron. Soc. Japan* **32**, 581–603 (1990).
10. Gardiner, L. T. & Noguchi, M. N-body simulations of the Small Magellanic Cloud and the Magellanic Stream. *Mon. Not. R. Astron. Soc.* **278**, 191–208 (1996).
11. Gardiner, L. T., Sawa, T. & Fujimoto, M. Numerical simulations of the Magellanic Clouds and the global gas distribution. *Mon. Not. R. Astron. Soc.* **266**, 567–582 (1994).
12. Lin, D. N. C., Jones, B. F. & Klemola, A. R. The motion of the Magellanic Clouds, origin of the Magellanic System, and the mass of the Milky Way. *Astrophys. J.* **439**, 652–671 (1995).
13. Davies, R. D. & Wright, A. E. A tidal origin for the Magellanic Stream. *Mon. Not. R. Astron. Soc.* **180**, 71–88 (1977).
14. Lin, D. N. C. & Lynden-Bell, D. Simulation of the Magellanic Stream to estimate the total mass of the Milky Way. *Mon. Not. R. Astron. Soc.* **181**, 59–81 (1977).
15. Guhathakurta, P. & Reitzel, D. B. in *In Galactic Halos: A UC Santa Cruz Workshop* (ed. Zaritsky, D.) 22–29 (ASP Conf. Ser., San Francisco, 1998).
16. Staveley-Smith, L. H I multibeam survey techniques. *Publ. Astron. Soc. Aus.* **14**, 111–116 (1997).
17. Barnes, D. G., Staveley-Smith, L., Ye, T. & Oosterloo, T. in *Astronomical Data Analysis Software and Systems (ADA) VII* (eds Albrecht, R., Hook, R. N. & Bushouse, H. A.) 89–92 (ASP Conf. Ser., San Francisco, 1998).
18. Mathewson, D. S. & Ford, V. L. in *Structure and Evolution of the Magellanic Clouds* (eds van den Bergh, S. & de Boer, K. S.) 125–136 (Kluwer, Dordrecht, 1984).
19. Morras, R. Fine structure in high velocity clouds near the South Celestial Pole. *Astron. Astrophys.* **115**, 249–252 (1982).
20. Wakker, B. P. Distribution and origin of high-velocity clouds. II—Statistical analysis of the whole-sky survey. *Astron. Astrophys.* **250**, 499–508 (1991).
21. Burton, W. B. in *Galactic and Extragalactic Radio Astronomy* (eds Verschuur, G. L. & Kellermann, K. I.) 295–358 (Springer, New York, 1998).
22. Brück, M. T. & Hawkins, M. R. S. An investigation of faint stars in a region of the Magellanic Stream. *Astron. Astrophys.* **124**, 216–222 (1983).
23. Osthimer, J. C., Majewski, S. R., Kunkel, W. E. & Johnston, K. V. A search for tidal star streams from the Magellanic Clouds. *Bull. Am. Astron. Soc.* **191**, 131.03 (1997).
24. Broeils, A. H. & van Woerden, H. A search for spiral galaxies with extended H I disks. *Astron. Astrophys. (suppl.)* **107**, 129–176 (1994).
25. Yun, M. S., Ho, P. T. P. & Lo, K. Y. A high-resolution image of atomic hydrogen in the M81 group of galaxies. *Nature* **372**, 530–532 (1994).
26. Kennicutt, R. C., Bresolin, E., Bomans, D. J., Bothun, G. D. & Thompson, I. B. Large scale structure of the ionized gas in the Magellanic Clouds. *Astron. J.* **109**, 594–604 (1995).
27. McGee, R. X. & Newton, L. M. H I profiles in the Bridge region of the Magellanic Clouds. *Proc. Astron. Soc. Aus.* **6**, 471–500 (1986).
28. Lu, L. *et al.* The metallicity and dust content of HVC 287.5+22.5+240: evidence for a Magellanic Clouds origin. *Astron. J.* **115**, 162–167 (1998).
29. Hartmann, D., Kalberla, P. M. W., Burton, W. B. & Mebold, U. Stray-radiation correction as applied to the Leiden/Dwingeloo survey of H I in the Galaxy. *Astron. Astrophys. (suppl.)* **119**, 115–151 (1996).

Acknowledgements. We acknowledge Warwick Wilson, Mal Sinclair and their teams at CSIRO for their engineering excellence and thank the AIPS++ team for their software support; we also thank E. de Blok, A. Green, S. Juraszek, M. Kesteven, R. Kraan-Korteweg, A. Schröder, R. Gooch and L. Gardiner.

Correspondence should be addressed to M.E.P. (e-mail: putman@mso.anu.edu.au).

Inevitability of a magnetic field in the Sun's radiative interior

D. O. Gough*† & M. E. McIntyre†

* Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

† Centre for Atmospheric Science at the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK

The gas in the convective outer layers of the Sun rotates faster at the equator than in the polar regions, yet deeper inside (in the radiative zone) the gas rotates almost uniformly^{1–3}. There is a thin transition layer between these zones, called the tachocline⁴. This structure has been measured seismologically^{1–3}, but no purely fluid-dynamical mechanism can explain its existence. Here we argue that a self-consistent model requires a large-scale magnetic field in the Sun's interior, as well as consideration of the Coriolis effects in the convection zone and in the tachocline. Turbulent stresses in the convection zone induce (through Coriolis effects) a meridional circulation, causing the gas from the convection zone to burrow downwards, thereby generating the horizontal and vertical shear that characterizes the tachocline. The interior magnetic field stops the burrowing, and confines the shear, as demanded by the observed structure of the tachocline. We outline a dynamical theory of the flow, from which we estimate a field strength of about 10^{-4} tesla just beneath the tachocline. An important test of this picture, after numerical refinement, will be quantitative consistency between the predicted and observed interior angular velocities.

Two quite different explanations for the near-uniform interior rotation of the Sun have been proposed. The first invokes quasi-horizontal shear-generated turbulence in the (stably stratified) tachocline, producing stresses that are predominantly horizontal^{4,5}. The stresses are assumed to act in the manner of a horizontal viscosity that leads to latitudinally invariant angular velocity beneath the tachocline. Therefore, if the rotation of the Sun were steady, the radiative zone beneath the tachocline would indeed be expected to rotate uniformly, though how radial shear would be suppressed in the presence of solar-wind spindown is not adequately explained. The equilibrium angular velocity of the radiative interior Ω_i is predicted to be $0.90\Omega_e$, where $\Omega_e \approx 2.9 \times 10^{-6} \text{ s}^{-1}$ is the equatorial angular velocity in the convective zone, and is somewhat less than that observed^{1–3}, namely $\Omega_i \approx 0.93\Omega_e$. The discrepancy seems significant.

As was recognized by Spiegel and Zahn⁴, the density stratification of the tachocline is far too strong to permit anything resembling three-dimensional turbulence⁶. In the presence of such stratification, the balance of forces is similar to that in the terrestrial stratosphere, about which there are a wealth of observations and strong theoretical arguments to show that horizontal turbulence controls the distribution of angular momentum in such a way as to drive the system away from, not towards, uniform rotation⁷. Meteorologists once called this negative viscosity⁸.

The second proposal^{9,10} is that shear-induced differential dissipation of prograde and retrograde gravity waves generated at the interface between the convective zone and the radiative interior acts in such a way as to suppress the shear in the radiative zone. But as already recognized in an earlier discussion of wave dissipation in the Sun¹¹, torques due to gravity waves do not behave in this way. Broadband gravity waves tend also to drive the system away from uniform rotation^{7,12}. There is a large body of theoretical studies coupled with observations of the Earth's atmosphere that bear on the issue of how gravity waves interact with differential rotation¹³. In

no known case in which there is a broad spectrum of angular phase speeds is the result of the interaction such as to drive the flow towards a state of uniform rotation. Torques due to acoustic waves appear to be far too weak¹⁴.

Our arguments, strongly reinforced by terrestrial stratospheric observations, suggest that no purely fluid-dynamical mechanism can explain the almost uniform interior rotation: hence the conclusion that shear is suppressed by a large-scale poloidal magnetic field in the interior. The idea that there might be such a field is not new¹⁵. What is new, however, is the inevitability of its existence. Without it, the tachocline would be much deeper than observed.

As a first step towards a realistic theory of the tachocline we adopt the simplest possible hypothesis: we ignore convective overshoot, shear-generated turbulence and wave-induced stresses, and any magnetic field that is generated in the convection zone. We consider the consequences of a downward-penetrating axisymmetric laminar

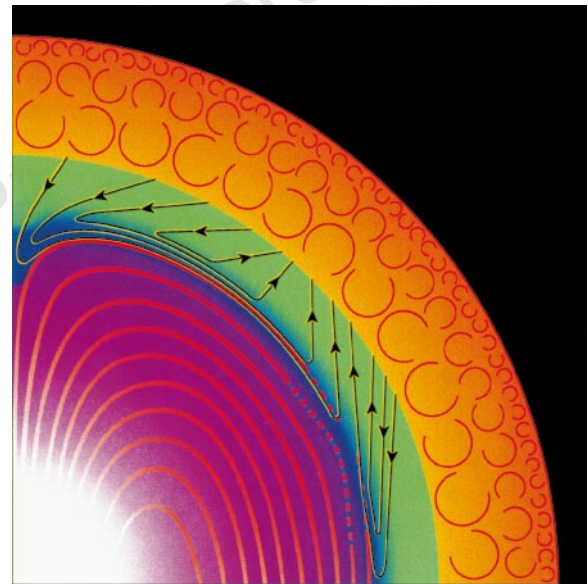


Figure 1 Schematic representation of a meridional quadrant of the Sun. The arrows represent the tachocline circulation, which follows surfaces S of constant specific angular momentum in the (green) body of the tachocline (whose thickness has been exaggerated by a factor of five), and is deflected by the magnetic field in the (blue) diffusive boundary layer (whose thickness has been exaggerated by a factor of 50). The inclination of the S -surfaces, which, owing to the exaggeration of the tachocline thickness are not drawn accurately, follow from the observation that the interior angular velocity Ω_i lies between the angular velocities at the equator and at the poles in the (orange) convection zone. Moreover, the centre of upwelling should be at a latitude of $\sim 30^\circ$ (where, incidentally, sunspots emerge at the start of a new cycle). We are unable to draw the return flow in the convection zone without knowledge of the Reynolds stresses; details in the mid-latitude upwelling region are also uncertain, obeying severely nonlinear dynamics, and may well be unsteady. The red lines represent the magnetic field in the (purple and white) radiative interior, which is assumed to be the dipole relic of a primordial field, arguably the most likely possibility (for simplicity, aligned with the rotation axis); we are unsure of the geometry of the field near the centre of upwelling, where the field lines are either dashed or absent. North-south asymmetry, as seen in the sunspot distributions observed in the Maunder minimum²⁴, may be related to the non-reversing interior dipole field. At the base of the tachocline the interior field vanishes on the rotation axis, where the magnetic boundary-layer theory suggests a singularity in the tachocline depth. The corresponding physical reality, which would again require nonlinear theory to describe it, would be relatively deep penetration of the tachocline circulation into and out of a 'polar pit', which might, conceivably, extend deep enough for lithium to be destroyed by nuclear reactions. The latest inversions of SOI seismic data³ suggest such a pit in the angular-velocity variation.

circulation forced by the turbulent stresses in the convection zone above. This type of circulation has been studied extensively in terrestrial stratospheric models^{16,17}. In the solar case downwelling occurs at both high and low latitudes, with upwelling in middle latitudes, as is illustrated schematically in Fig. 1. The interior magnetic field stops the downwelling branches of the forced circulation from burrowing more and more deeply as time goes on. There is a ‘tachopause’ in the form of a thin magnetic boundary layer within which the angular-velocity gradient is reduced to zero. The principal balance of forces is summarized below. By hypothesis, there is no azimuthal force on the downwelling flow above the magnetic boundary layer in the body of the tachocline; therefore angular momentum is conserved. In a quasi-steady state, the flow in the body of the tachocline is along surfaces S of constant specific angular momentum, which from helioseismology are known to slant qualitatively as depicted in the figure.

Only where there is upwelling is there an opportunity for the interior magnetic field to penetrate into the body of the tachocline and thence into the convection zone, where it will be subject to reconnection. The magnetohydrodynamic problem in the upwelling region is severely nonlinear, and its details formidably complicated, perhaps modifying angular-momentum conservation via Lorentz forces and reshaping the upwelling region, which mass conservation demands must still exist.

The dynamics implied by this picture relates tachocline depth to interior field strength. First, in the body of the tachocline, hydrostatic and cyclostrophic balance relates the shear of angular velocity to the temperature field T , or, more precisely, to the aspherical temperature anomaly $\bar{T}$ induced by the downwelling (warm, compressing) or upwelling (cold, expanding) flow, in a manner that depends on the depth Δ of the tachocline. It proves convenient to introduce spherical polar coordinates (r, θ, ϕ) with respect to the rotation axis, in a frame rotating with the angular velocity Ω_i of the radiative interior. In this frame the velocity $\mathbf{u} = (u, v, s\bar{\Omega})$, where $s = r\sin\theta$, and the angular velocity anomaly $\bar{\Omega}$ depends on r and θ and describes the tachocline shear. The balance is then expressed by the zonal component of the steady axisymmetric inviscid vorticity equation, yielding an appropriately generalized ‘thermal-wind’¹³ (more aptly, ‘thermal-shear’) equation, which, assuming the perfect-gas law, is

$$2\Omega_i s \left(\frac{\partial \bar{\Omega}}{\partial z} \right)_s \approx \frac{g}{rT} \left(\frac{\partial \bar{T}}{\partial \theta} \right)_r \quad (1)$$

where $z = r\cos\theta$ and g is the gravitational acceleration.

Second, the thermal-energy equation expresses a balance between radiative diffusion and advection of the background stratification:

$$\frac{N^2 T u}{g} \approx \frac{1}{\rho c_p r^2} \frac{\partial}{\partial r} \left(r^2 K \frac{\partial \bar{T}}{\partial r} \right) \quad (2)$$

the partial derivatives being taken at constant colatitude θ , where N is the buoyancy frequency, ρ the mass density, c_p the specific heat at constant pressure and K the radiative conductivity. In the downwelling zones, these equations are to be solved subject to the conditions that the flow is along S and that $\bar{\Omega}$ and $\bar{T}$ match smoothly to the corresponding variables in the radiative interior and the convection zone. In particular, $\bar{\Omega} \rightarrow 0$ and $\bar{T} \rightarrow 0$ towards the base of the tachocline.

Third, the downwelling in the polar and the sub-equatorial regions is deflected by the interior magnetic field towards the mid-latitude upwelling region. There is a thin magnetic boundary layer in which aspherical deviations vary on a vertical length scale δ ; we confirm *a posteriori* that $\delta \ll \Delta$. This boundary layer separates the body of the tachocline from what the model assumes to be a rigidly rotating radiative interior. The location of the boundary layer is determined by a balance between downward advection

and upward diffusion of the large-scale, essentially horizontal, poloidal magnetic field: therefore $\eta/\delta \approx |u|$, where η is the magnetic diffusivity.

Equations (1) and (2) continue to be satisfied within the boundary layer (again confirmed *a posteriori*), but now the angular-momentum balance is modified by the Lorentz force. The force balance is expressed by the zonal component of the momentum equation:

$$2\Omega_i v \cos\theta \approx \frac{B_0}{\mu_0 \rho r \sin\theta} \frac{\partial}{\partial \theta} (B_\phi \sin\theta) \quad (3)$$

where μ_0 is the vacuum permeability and B_0 is the poloidal magnetic field, which is diffusing upwards from the radiative interior and is approximated here as being purely horizontal; B_ϕ is the zonal component of the magnetic field, resulting from the twisting and stretching of the poloidal field by the differential rotation according to the zonal component of the induction equation which, near the bottom of the magnetic boundary layer where the tachocline perturbation is small, is

$$-B_0 \sin\theta \frac{\partial \bar{\Omega}}{\partial \theta} \approx \eta \frac{\partial^2 B_\phi}{\partial r^2} \quad (4)$$

Consistent with the horizontal- B_0 approximation, we have ignored latitudinal variations of $B_0 \sin\theta$. Equations (1)–(4), coupled with the continuity equation $\text{div}(\rho \mathbf{u}) = 0$, can then be reduced to a single linear equation for $\bar{T}$. The latitudinal variation of $\bar{T}$ is oscillatory on a scale comparable to the radius $r_c \approx 0.7R_\odot$ of the base of the convection zone (here $R_\odot$ is the solar radius). Accordingly, we may analyse the boundary layer locally, introducing a characteristic horizontal wavenumber L/r_c and setting $\partial/\partial\theta = iL$. The boundary-layer equation then reduces to

$$\frac{\partial^6 \bar{T}}{\partial r^6} - \delta^{-6} \bar{T} \approx 0 \quad (5)$$

in which the boundary-layer thickness scale is given by

$$\delta = \left(\frac{2\mu_0 \rho \eta \kappa \Omega_i^2}{B_0^2 r_c^2 N^2 L^4} \right)^{1/6} r_c \quad (6)$$

where $\kappa = K/\rho c_p$ is the radiative diffusivity. In obtaining equation (5) it was possible to neglect the variation of the properties of the unperturbed state because the magnetic boundary layer is thin. The equation has three solutions that decay with depth, all on scales of order $\delta \ll \Delta$, a combination of which must be matched to the solution in the body of the tachocline.

Near the bottom of the magnetic boundary layer the linearization leading to equations (4) and (5) is valid, and shows how the tachocline circulation is prevented by the interior magnetic field from burrowing more deeply. But as the body of the tachocline is approached, with $|u|$ rising to values $\approx \eta/\delta$, advection of B_ϕ by the meridional circulation (not included in equation (4)), and advective modification of B_0 , both become important, and the boundary-layer equations become nonlinear. We have not yet solved these equations, but we can make an estimate of a typical value of $|B_0|$ from scaling arguments. We expect the latitudinal variation of $\bar{T}$ to be similar to that of the spherical harmonic $P_4(\cos\theta)$, and so we adopt the estimate $L \approx 4.5$. We take $\partial/\partial r \approx \beta/\Delta$ for equation (2) in the body of the tachocline, where $\beta \approx \pi$, and we represent the jump in $\bar{\Omega}$ across the tachocline as $\alpha\Omega_i$, and neglect its jump across the boundary layer. Then equations (1) and (2) imply $\bar{T}/T \approx \alpha r_c^2 \Omega_i^2 / g L \Delta$ and $u \approx \beta^2 \kappa g \bar{T} / \Delta^2 N^2 T$, where we have taken $\sin\theta \approx \cos\theta \approx 2^{-1/2}$. We expect these estimates to apply everywhere except near the equator where the S -surfaces fail to traverse the tachocline, and near the poles, where our boundary-layer theory breaks down.

Together with the magnetic advection–diffusion balance,

$|u| \approx \eta/\delta$, these estimates imply:

$$\delta \approx \frac{L}{\alpha\beta^2} \frac{\eta}{\kappa} \left(\frac{N}{\Omega_i}\right)^2 \left(\frac{\Delta}{r_c}\right)^3 r_c \quad (7)$$

Equating (6) and (7) relates the local Alfvén speed $|B_0|/\sqrt{\mu_0\rho}$ to the global diffusion speed $\sqrt{\kappa\eta/r_c}$:

$$\frac{|B_0|}{\sqrt{\mu_0\rho}} \approx \frac{\sqrt{2}\alpha^3\beta^6}{L^5} \left(\frac{\kappa}{\eta}\right)^3 \left(\frac{\Omega_i}{N}\right)^7 \left(\frac{r_c}{\Delta}\right)^9 \frac{\sqrt{\kappa\eta}}{r_c} \quad (8)$$

The sign of B_0 is immaterial. Substituting the characteristic values $\alpha \approx 0.07$, $\kappa/\eta \approx 1.4 \times 10^4$, $\Omega_i/N \approx 3 \times 10^{-3}$, $\sqrt{\kappa\eta}/r_c \approx 2 \times 10^{-8} \text{ m s}^{-1}$, $\rho \approx 2 \times 10^2 \text{ kg m}^{-3}$, $\mu_0 = 4\pi \times 10^{-7} \text{ T}^2 \text{ Pa}^{-1}$, yields:

$$|B_0| \approx 5 \times 10^{-19} \left(\frac{r_c}{\Delta}\right)^9 \text{ T} \quad (9)$$

This shows how the tachocline thickness Δ depends on the interior field B_0 at the base of the tachocline.

In current standard models of the Sun, helium settles under gravity on a timescale $\sim 10^{10}$ yr, resulting in a negative abundance gradient throughout the radiative interior. Our model implies that material in the tachocline is recirculated into the convection zone in the ventilation time τ_v , which we shall find to be $\ll 10^{10}$ yr, implying in turn that the helium abundance is uniform down to the tachopause. The upshot is that the mean molecular mass of the tachocline is lower than that in the standard models, with a consequent sound-speed anomaly immediately beneath the convection zone. This phenomenon is observed in the most recent structure inversions of p-mode frequencies obtained from GONG¹⁸ and SOHO/SOI². Seismic calibration of our model¹⁹ yields $\Delta \approx 0.018R_\odot \approx 0.026r_c$. Hence, from equation (9), it follows that the intensity of the magnetic field immediately underlying the magnetic boundary layer is $|B_0| \approx 10^{-4} \text{ T}$. (An earlier report²⁰ of a much more intense, toroidal, field inferred by seismic inversion could be a misinterpretation of just such a sound-speed anomaly, from which a magnetic field cannot be distinguished by frequency inversion alone²¹).

From equation (7) one can now estimate the characteristic scale of the magnetic boundary layer to be $\delta \approx 10^{-3} r_c \approx 4 \times 10^{-2} \Delta$, confirming that $\delta \ll \Delta$. It can be confirmed also that viscous stresses are wholly negligible. Furthermore, using $g \approx 530 \text{ m s}^{-2}$, we find $\tilde{T}/T \approx 4 \times 10^{-6}$ and, using $\eta \approx 0.07 \text{ m}^2 \text{ s}^{-1}$, we find $u \approx \eta/\delta \approx 10^{-7} \text{ m s}^{-1}$, from which one obtains the characteristic tachocline ventilation time $\tau_v \approx \Delta/u \approx 3 \times 10^6 \text{ yr}$. If the magnetic field B_i deep in the radiative interior is the remnant of a primordial field, it is presumably stronger than B_0 , conceivably by a factor as great as $R_\odot/2\delta$, in which case $B_i \approx 0.1 \text{ T}$, which is rather less than a previous rough theoretical estimate of the strength of the protosolar field²².

The theory that we have outlined is only a preliminary analysis based on linearized boundary-layer equations, which is enough to establish the smallness of the scale δ and, most importantly, the ability of the magnetic boundary layer to stop the downwelling, and suppress the shear.

We acknowledge that our simple model does not take into account several dynamical processes believed to be present in the Sun, such as convective overshoot, shear-generated turbulence and wave-induced stresses. Moreover, because the model assumes a steady flow, we assume there to be no effect from the spindown, and no rectified trace of the 22-year solar-cycle variations that might be carried in the downwelling to affect the tachocline circulation. We estimate that, although any of these processes might actually modify somewhat the magnitude and geometry of the tachocline flow, they are unlikely to alter the model in any serious qualitative respect. The same is true of the flow within $\sim 15^\circ$ of the equator,

where the leading-order dynamics is different because the S-surfaces intersecting the base of the convection zone do not extend to the base of the tachocline.

The only magnetic field in the model is a remnant of the primordial field in the radiative core. Any field from a putative dynamo in the convection zone could be 'dredged' into the tachocline by the meridional flow, and thereby influence the dynamics; but it seems unlikely that the rapidly oscillating field associated with the solar cycle would contribute significantly to the dynamics in the radiative zone²³, particularly in view of the 10^6 yr tachocline ventilation time τ_v .

Seismic calibration of tachocline models having different depths Δ , using our expectation of a near-discontinuity at the thin magnetic boundary layer, estimates Δ much more precisely than the resolving power of any raw inversion. This is fortunate, because the estimate (from equation (9)) for $|B_0|$ varies inversely with the ninth power of Δ , and is therefore sensitive to measurement errors. The calibration reproduces the helioseismic sound-speed inversions extremely well¹⁹.

The picture we have arrived at provides a natural (and simplest possible), consistent view of the dynamical coupling between the convective and radiative zones, and points towards a comprehensive and realistic theory of the interior magnetic field and of spindown. Such a theory would have far-reaching implications for helioseismological inversion of the differential rotation, because Ferraro's law of isorotation (Ω constant on magnetic field lines¹⁵) would have to apply, under the simplest hypothesis of no interior torsional oscillation. Moreover, if the nonlinear boundary-layer and tachocline equations were to be solved—a formidable but probably tractable problem—it would be possible to deduce quantitatively the magnetic field at the base of the tachocline. $\square$

Received 13 March; accepted 21 May 1998.

- Thompson, M. J. *et al.* Differential rotation and dynamics of the solar interior. *Science* **272**, 1400–1405 (1996).
- Kosovichev, A. G. *et al.* Structure and rotation of the solar interior: initial results from the MDI medium-*l* program. *Sol. Phys.* **170**, 43–61 (1997).
- Schou, J. *et al.* Helioseismic studies with SOI-MDI of differential rotation in the solar envelope. *Astrophys. J.* (in the press).
- Spiegel, E. A. & Zahn, J.-P. The solar tachocline. *Astron. Astrophys.* **265**, 106–114 (1992).
- Elliott, J. R. Aspects of the solar tachocline. *Astron. Astrophys.* **327**, 1222–1229 (1997).
- Turner, J. S. *Buoyancy Effects in Fluids* (Cambridge Univ. Press, 1973).
- McIntyre, M. E. in *The Solar Engine and its Influence on Terrestrial Atmosphere and Climate* (ed. Nesme-Ribes, E.) 293–320 (NATO ASI 125, Springer, Berlin, 1994).
- Starr, V. P. *Physics of Negative-viscosity Phenomena* (McGraw Hill, New York, 1968).
- Kumar, P. & Quataert, E. J. Angular momentum transport by gravity waves and its effect on the rotation of the solar interior. *Astrophys. J.* **475**, L133–L136 (1997).
- Zahn, J.-P., Talon, S. & Matias, J. Angular momentum transport by internal waves in the solar interior. *Astron. Astrophys.* **322**, 320–328 (1997).
- Gough, D. O. in *The Energy Balance and Hydrodynamics of the Solar Chromosphere and Corona* (eds Bonnet, R.-M. & Delache, P.) 3–36 (IAU Colloq 36, de Bussac, Clermont-Ferrand, 1977).
- Plumb, R. A. & McEwan, A. D. The instability of a fixed standing wave in a viscous stratified fluid: a laboratory analogue of the quasi-biennial oscillation. *J. Atmos. Sci.* **35**, 1827–1839 (1978).
- Andrews, D. G., Holton, J. R. & Leovy, C. B. *Middle Atmosphere Dynamics* (Academic, New York, 1987).
- Gough, D. O. in *Solar-terrestrial Relationships and the Earth Environment in the Last Millennium* (ed. Castagnoli-Cini) 90–142 (Soc. Italiana Fisica, Bologna, 1988).
- Cowling, T. G. *Magnetohydrodynamics* (Interscience, New York, 1957).
- Haynes, P. H., McIntyre, M. E., Shepherd, T. G. Reply to Comments by J. Egger on 'On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces'. *J. Atmos. Sci.* **53**, 2105–2107 (1996).
- Haynes, P. H., Marks, C. J., McIntyre, M. E., Shepherd, T. G. & Shine, K. P. On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.* **48**, 651–678 (1991).
- Gough, D. O. *et al.* The seismic structure of the sun. *Science* **272**, 1296–1299 (1996).
- Elliott, J. R. & Gough, D. O. The thickness of the solar tachocline. *Astrophys. J.* (submitted).
- Dziembowski, W. A. & Goode, P. R. The toroidal magnetic field inside the sun. *Astrophys. J.* **347**, 540–550 (1989).
- Zweibel, E. G. & Gough, D. O. in *Proc. IV SOHO Workshop: Helioseismology* (eds Hoeksema, J. T., Domingo, V., Fleck, B. & Battrick, B.) 47–48 (SP-376, Vol. 2, ESA, Noordwijk, 1995).
- Gough, D. O. On possible origins of relatively short-term variations in the solar structure. *Phil. Trans. R. Soc. A* **330**, 627–640 (1990).
- Garaud, P. Propagation of a dynamo field in the radiative zone of the sun. *Mon. Not. R. Astron. Soc.* (submitted).
- Nesme-Ribes, E., Sokoloff, D., Ribes, J. C. & Kremlovsky, M. in *The Solar Engine and its Influence on Terrestrial Atmosphere and Climate* (ed. Nesme-Ribes, E.) 71–97 (NATO ASI I 25, Springer, Berlin, 1994).

Correspondence and requests for materials should be addressed to D.O.G. (e-mail: douglas@ast.cam.ac.uk).

Archetypal energy landscapes

David J. Wales, Mark A. Miller & Tiffany R. Walsh

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, UK

Energy landscapes hold the key to understanding a wide range of molecular phenomena. The problem of how a denatured protein re-folds to its active state (Levinthal's paradox¹) has been addressed in terms of the underlying energy landscape^{2–7}, as has the widely used 'strong' and 'fragile' classification of liquids^{8,9}. Here we show how three archetypal energy landscapes for clusters of atoms or molecules can be characterized in terms of the disconnectivity graphs¹⁰ of their energy minima—that is, in terms of the pathways that connect minima at different threshold energies. First we consider a cluster of 38 Lennard–Jones particles, whose energy landscape is a 'double funnel' on which relaxation to the global minimum is diverted into a set of competing structures. Then we characterize the energy landscape associated with the annealing of C₆₀ cages to buckminsterfullerene, and show that it provides experimentally accessible clues to the relaxation pathway. Finally we show a very different landscape morphology, that of a model water cluster (H₂O)₂₀, and show how it exhibits features expected for a 'strong' liquid. These three examples do not exhaust the possibilities, and might constitute substructures of still more complex landscapes.

We use disconnectivity graphs¹⁰ to visualize the energy landscapes, based upon samples of pathways linking local minima via transition states. At any given total energy we elucidate which of the local minima in our sample are connected by pathways that lie

below the energy threshold. At finite energy the minima are divided into disconnected sets of mutually accessible structures, separated by insurmountable barriers. The resulting graphs are clearest when we present the results as the total energy increases in regular steps along the vertical axis. Each line begins from a different local minimum at the bottom of the corresponding well. A node joins lines at the lowest energy for which the minima become mutually accessible. We are free to choose the horizontal displacements to give the most helpful representation of the resulting graph.

Graphs such as these, which are connected but contain no cycles, are known as 'trees' for reasons which should be clear from Fig. 1. The gentle 'funnel' with high barriers in Fig. 1a produces a graph which looks like a weeping willow. The graph for the efficient funnel with lower barriers in Fig. 1b reminds us of a palm tree, whilst the graph for the rough landscape in Fig. 1c resembles a banyan tree with its branches planted into the ground.

We first focus on a cluster of 38 atoms bound by the Lennard–Jones potential, represented here by (LJ)₃₈. The lowest-energy icosahedral minimum lies significantly above the truncated octahedron¹¹, and the two minima are well separated in configuration space. Relaxation to the global minimum is hampered because the vast majority of local minima are associated with the liquid-like state of the cluster and have polytetrahedral character¹¹. Minima based upon icosahedra also have polytetrahedral character, but the truncated octahedron is a fragment of close-packed lattice.

Hence the (LJ)₃₈ cluster provides a pattern for the 'double funnel'

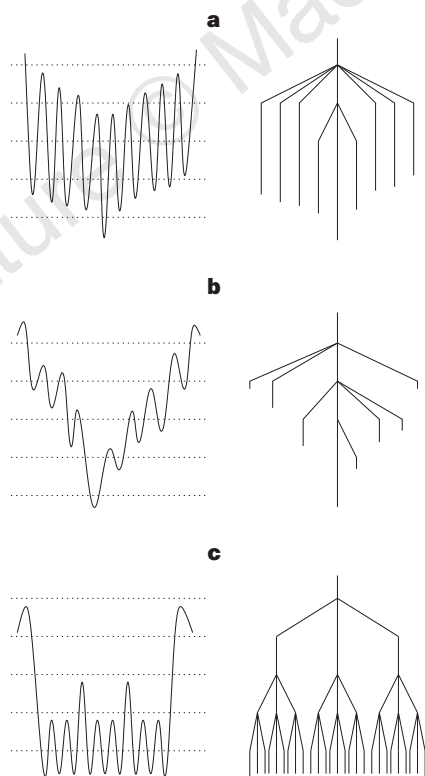


Figure 1 Pictorial correspondence between the potential-energy surface and the disconnectivity graph for three different energy landscapes, following Becker and Karplus¹⁰. **a**, The 'weeping willow' results from a gentle funnel with large barriers. **b**, The 'palm tree' results from a steeper funnel with lower barriers. **c**, The 'banyan tree' results from a rough landscape.

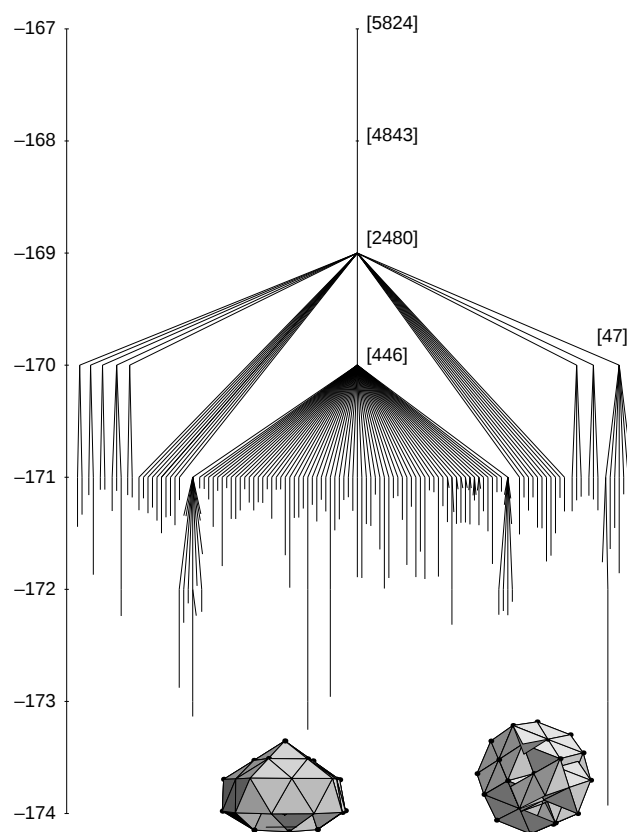


Figure 2 Disconnectivity graph for the double-funnel surface of (LJ)₃₈. The energy on the vertical axis is measured in terms of the pair well depth. The numbers in square brackets indicate the number of minima represented by the adjacent node; only branches leading to the lowest 150 minima are shown explicitly.

surface, and understanding the dynamics of this system should provide insight into landscapes which support competing morphologies. Our sample of stationary points for $(\text{LJ})_{38}$ contains 6,000 minima and 8,633 transition states, along with the corresponding pathways. The construction of such databases has been described before^{2,12} and the resulting graph is shown in Fig. 2. For clarity we have removed branches that do not lead to the lowest 150 minima. The graph shows that of these minima, most lie in the region of configuration space associated with icosahedra, rather than with the global minimum. Hence, relaxation to the global minimum is diverted by the presence of a competing morphology associated with a much larger region of configuration space and higher entropy.

We next turn to C_{60} . A number of studies^{13–21} have proposed mechanistic schemes to explain the formation of buckminsterfullerene, but simulations of fullerene nucleation usually produce defective structures^{22–27}. The exceptions are a calculation starting from the adjacent local minimum²⁸ and a study where the dynamics appear to be accelerated because the potential employed has rather low barriers²⁹. For potentials which support realistic

barriers, the timescale for fullerene annealing lies beyond the reach of traditional simulation techniques. The present calculations employed a semi-empirical, quantum-mechanical tight-binding potential³⁰.

We have deduced the disconnectivity graph for C_{60} using the most likely annealing mechanism, assuming that a C_{60} molecule with 12 pentagonal and 20 hexagonal faces has already been reached somehow, and neglecting fullerenes with square or heptagonal faces. This approach seems reasonable, as both the 'pentagon road' and 'hexagon road' growth mechanisms entail annealing after the formation of higher-energy C_{60} fullerene isomers^{14–16}.

Excluding enantiomers, there are 1,812 different isomers for C_{60} with 12 pentagonal and 20 hexagonal faces, and following Austin *et al.*³¹ we have performed geometry optimizations for all these minima. Austin *et al.*³² have also investigated the connectivity of the same set of minima under a generic rearrangement mechanism proposed by Stone and Wales³³. This 'pyraclyene' or 'Stone–Wales' mechanism provides a means for pentagonal and hexagonal rings to migrate over the surface of the cluster. The

Figure 3 The 'pyraclyene' or 'Stone–Wales' rearrangement of C_{60} . Buckminsterfullerene, left, is transformed via the transition state, centre, to give the new C_{2v} symmetry minimum, right.

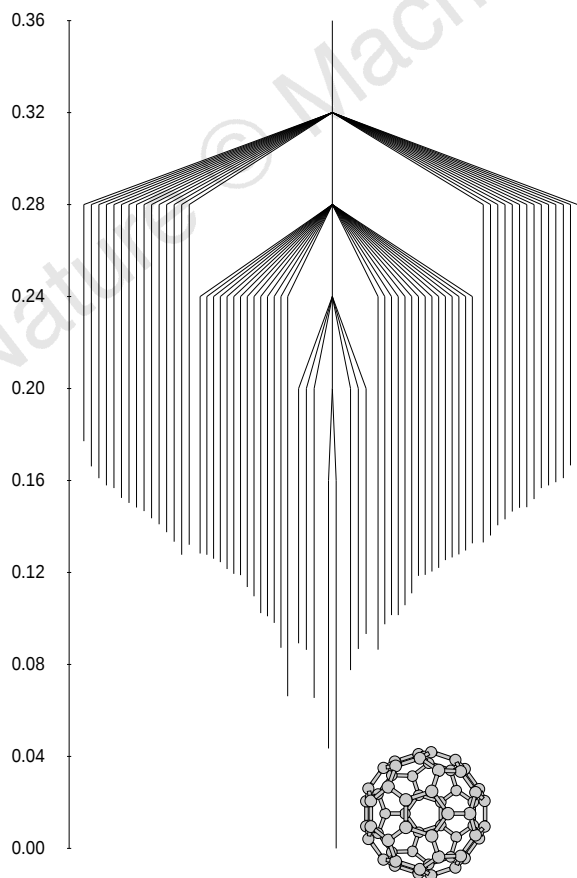
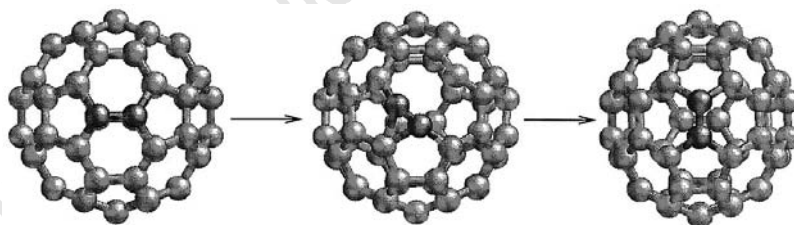


Figure 4 Disconnectivity graph for minima and transition states in the five lowest Stone–Wales stacks of C_{60} . Energy is in hartrees relative to the global minimum buckminsterfullerene structure.

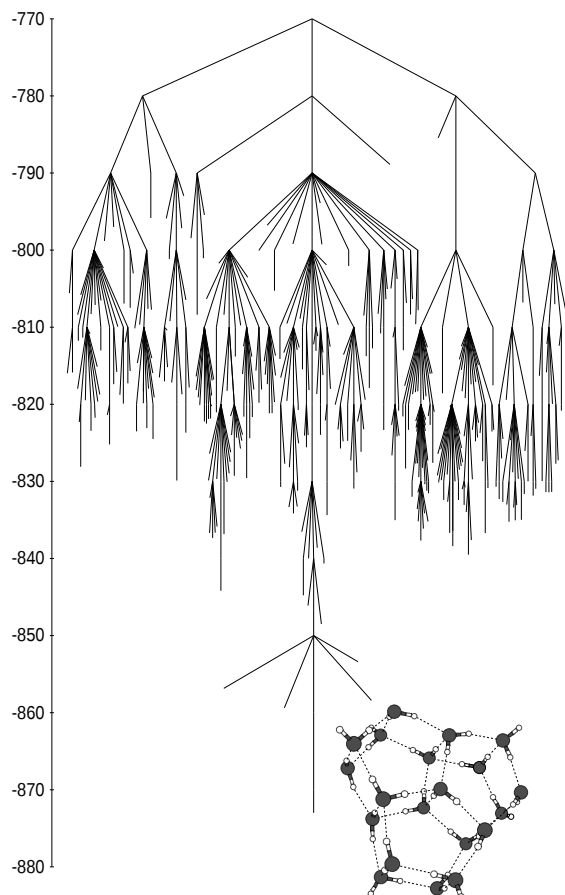


Figure 5 Disconnectivity graph for a sample of minima and transition states around the global minimum of the TIP4P $(\text{H}_2\text{O})_{20}$ cluster. Energy is in kJ mol^{-1} .

transition state that we have found for this process appears to agree quite well with the 'sp³ intermediate' reported by Scuseria and co-workers^{18,34}, although our barrier of 500 kJ mol⁻¹ is somewhat lower than theirs and closer to that obtained by Yi and Bernholc²⁷ for C₆₀ and by Baker and Fowler³⁵ for C₂₈. Our calculated pathway linking the icosahedral isomer to the next lowest isomer is shown in Fig. 3.

Austin *et al.*³² determined that the largest subset of the 1,812 C₆₀ isomers which are mutually accessible via Stone–Wales rearrangements contains 1,710 minima, one of which is buckminsterfullerene. In fact, the minima can be grouped into 'stacks' according to how many Stone–Wales processes are required to convert them into buckminsterfullerene³². The first stack contains buckminsterfullerene itself, and the second stack contains the adjacent minimum illustrated in Fig. 3. Here we have calculated all the Stone–Wales rearrangement mechanisms up to, and including, those linking the minima in stack seven. This set includes 547 pathways, and 197 of the 1,812 local minima. We include minima only up to stack five in Fig. 4 for clarity. This figure clearly corresponds to the 'weeping willow' pattern of Fig. 1b, that is, a gentle funnel with higher barriers.

Using the above graph we have applied a master equation approach to examine the relaxation dynamics of these C₆₀ isomers. A high temperature is required to produce an appreciable population of buckminsterfullerene on the experimentally observed timescales, in good agreement with the experimental observation that high-temperature annealing is needed to produce significant yields of this structure^{16,36,37}. These results reflect the nature of the energy landscape revealed in Fig. 4, namely the presence of a funnel leading to the global minimum but with relatively large barriers.

Our final example is the energy landscape of a (H₂O)₂₀ cluster described by the TIP4P rigid-molecule effective pair potential³⁸. Although this potential was designed to reproduce the properties of bulk water it can provide a useful starting point for cluster studies, despite the neglect of many-body forces. Here our objective is to explain Angell's classification of water as a 'strong' liquid in terms of the energy landscape^{8,9}, and a simple model with the right anisotropic form should suffice. 'Fragile' liquids have structures which change significantly with temperature, whilst 'strong' liquids often have open network structures which are preserved over relatively wide temperature ranges, and are characterized by viscosities which exhibit Arrhenius temperature dependence⁸, that is, they scale as exp(constant/*T*) with temperature, *T*.

The global minimum for TIP4P (H₂O)₂₀ can be described in terms of three fused pentagonal prisms³⁹. The graph for a sample of minima in the vicinity of this structure is shown in Fig. 5. Apart from the lowest-lying isomers, the graph resembles the 'banyan tree' shown in Fig. 1c. This pattern arises from a hierarchical structure: sets of minima are progressively connected by transition states of increasing energy, and the resulting structure is similar to that expected for a fractal¹⁰. Despite the wide range of transition-state energies, almost all the barrier heights are less than 20 kJ mol⁻¹, the energy of a typical hydrogen bond, and fall within a relatively narrow range. Nevertheless, the corresponding rearrangement mechanisms include both cooperative and localized processes, and the barriers are relatively large compared to the available thermal energy at temperatures where water can exist as a liquid. Global optimization for such a system is difficult because the path between a randomly chosen structure and the global minimum is likely to involve more local minima, including some of relatively high energy³⁹.

We have found that smaller water clusters have considerably fewer minima on the potential-energy surface than atomic Lennard–Jones clusters with the same number of degrees of freedom.

We attribute this result to the longer range and anisotropy of the forces between the water molecules. The same is probably true for larger water clusters, again in agreement with previous descriptions of 'strong' liquids⁴⁰. □

Received 20 April; accepted 8 June 1998.

- Levinthal, C. In *Mössbauer Spectroscopy in Biological Systems, Proceedings of a Meeting Held at Allerton House, Monticello, Illinois* (eds DeBrunner, J. T. P. & Muck, E.) 22–24 (Univ. Illinois Press, Urbana, 1969).
- Ball, K. D. *et al.* From topography to dynamics on multidimensional potential energy surfaces of atomic clusters. *Science* **271**, 963–966 (1996).
- Wales, D. J. Tales from topographic potential surfaces. *Science* **271**, 925–929 (1996).
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnel, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).
- Doye, J. P. K. & Wales, D. J. On potential energy surfaces and relaxation to the global minimum. *J. Chem. Phys.* **105**, 8428–8445 (1996).
- Wolynes, P. G. Symmetry and the energy landscapes of biomolecules. *Proc. Natl Acad. Sci. USA* **93**, 14249–14255 (1996).
- Berry, R. S., Elmaci, N., Rose, J. P. & Vekhter, B. Linking topography of its potential surface with the dynamics of folding of a protein model. *Proc. Natl Acad. Sci. USA* **94**, 9520–9524 (1997).
- Angell, C. A. Formation of glasses from liquids and biopolymers. *Science* **267**, 1924–1935 (1995).
- Stillinger, F. H. A topographic view of supercooled liquids and glass formation. *Science* **267**, 1935–1939 (1995).
- Becker, O. M. & Karplus, M. The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics. *J. Chem. Phys.* **106**, 1495–1517 (1997).
- Wales, D. J. & Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of Lennard–Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111–5116 (1997).
- Doye, J. P. K. & Wales, D. J. Surveying a potential energy surface by eigenvector-following. *Z. Phys. D* **40**, 194–197 (1997).
- Kroto, H. W. Space, stars, C₆₀, and soot. *Science* **242**, 1139–1135 (1988).
- Heath, J. R., O'Brien, S. C., Curl, R. F., Kroto, H. W. & Smalley, R. E. Carbon condensation. *Comments Cond. Mat. Phys.* **13**, 119–141 (1987).
- Heath, J. R. In *Fullerenes—Synthesis, Properties, and Chemistry of Large Carbon Clusters* (eds Hammond, G. S. & Kuck, V. J.) 1–22 (ACS Symp. Ser. 481, Am. Chem. Soc., Washington DC, 1992).
- Smalley, R. E. Self-assembly of the fullerenes. *Acc. Chem. Res.* **25**, 98–105 (1992).
- Endo, M. & Kroto, H. W. Formation of carbon nanofibers. *J. Phys. Chem.* **96**, 6941–6944 (1992).
- Murry, R. L., Strout, D. L., Odum, G. K. & Scuseria, G. E. Role of sp³ carbon and 7-membered rings in fullerene annealing and fragmentation. *Nature* **366**, 665–667 (1993).
- Manopoulos, D. E. & Fowler, P. W. In *The Far-reaching Impact of the Discovery of C₆₀* (ed. Andreoni, W.) 51–69 (NATO ASI Ser. E **316**, Kluwer, Dordrecht, 1993).
- Strout, D. L. & Scuseria, G. E. A cycloaddition model for fullerene formation. *J. Phys. Chem.* **100**, 6492–6498 (1996).
- Marcos, P. A., López, M. J., Rubio, A. & Alonso, J. A. Thermal road for fullerene annealing. *Chem. Phys. Lett.* **273**, 367–370 (1997).
- Ballone, P. & Milani, P. Simulated annealing of carbon clusters. *Phys. Rev. B* **42**, 3201–3204 (1990).
- Chelikowsky, J. R. Nucleation of C₆₀ clusters. *Phys. Rev. Lett.* **67**, 2970–2973 (1991).
- Chelikowsky, J. R. Formation of C₆₀ clusters via Langevin molecular-dynamics. *Phys. Rev. B* **45**, 12062–12070 (1992).
- Jing, X. D. & Chelikowsky, J. R. Nucleation of carbon clusters via an accretion model. *Phys. Rev. B* **46**, 5028–5031 (1992).
- Wang, C. Z., Xu, C. H., Chan, C. T. & Ho, K. M. Disintegration and formation of C₆₀. *J. Phys. Chem.* **96**, 3563–3565 (1992).
- Yi, J. Y. & Bernholc, J. Reactivity, stability, and formation of fullerenes. *Phys. Rev. B* **48**, 5724–5727 (1993).
- Xu, C. H. & Scuseria, G. E. Tight-binding molecular-dynamics simulations of fullerene annealing and fragmentation. *Phys. Rev. Lett.* **72**, 669–672 (1994).
- Maruyama, S. & Yamaguchi, Y. A molecular dynamics demonstration of annealing to a perfect C₆₀ structure. *Chem. Phys. Lett.* **286**, 343–349 (1998).
- Porezag, D., Frauenheim, T., Seifert, G. & Kaschner, R. Construction of tight-binding-like potentials on the basis of density-functional theory—application to carbon. *Phys. Rev. B* **51**, 12947–12957 (1995).
- Austin, S. J., Fowler, P. W., Manopoulos, D. E., Orlandi, G. & Zerbetto, F. Structural motifs and the stability of fullerenes. *J. Phys. Chem.* **99**, 8076–8081 (1995).
- Austin, S. J., Fowler, P. W., Manopoulos, D. E. & Zerbetto, F. The Stone–Wales map for C₆₀. *Chem. Phys. Lett.* **235**, 146–151 (1995).
- Stone, A. J. & Wales, D. J. Theoretical studies of icosahedral C₆₀ and some related species. *Chem. Phys. Lett.* **128**, 501–503 (1986).
- Scuseria, G. E. *Ab-initio* calculations of fullerenes. *Science* **271**, 942–945 (1996).
- Baker, J. & Fowler, P. W. Energetics of the Stone–Wales pyracene transformation. *J. Chem. Soc., Perkin Trans. 2*, 1665–1666 (1992).
- Zhang, Q. L. *et al.* Reactivity of large carbon clusters—spheroidal carbon shells and their possible relevance to the formation and morphology of soot. *J. Phys. Chem.* **90**, 525–528 (1986).
- Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. Solid C₆₀—a new form of carbon. *Nature* **347**, 354–358 (1990).
- Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 24. Revised TIPS for simulations of liquid water and aqueous-solutions. *J. Chem. Phys.* **77**, 4156–4163 (1982).
- Wales, D. J. & Hodges, M. P. Global minima of water clusters (H₂O)_n, *n* ≤ 21, described by an empirical potential. *Chem. Phys. Lett.* **286**, 65–72 (1998).
- Angell, C. A. Relaxation in liquids, polymers and plastic crystals—strong fragile patterns and problems. *J. Non-Cryst. Solids* **131–133**, 13–31 (1991).

Acknowledgements. We thank J. P. K. Doye and A. J. Stone for discussions. D.J.W., M.A.M. and T.R.W. thank the Royal Society, the EPSRC and the Cambridge Commonwealth Trust, respectively, for financial support.

Correspondence and requests for materials should be addressed to D.J.W. (e-mail: dw34@cus.cam.ac.uk).

Electron–electron correlations in carbon nanotubes

Sander J. Tans*, Michel H. Devoret*†, Remco J. A. Groeneveld* & Cees Dekker*

* Department of Applied Physics and DIMES, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

† Service de Physique de l'Etat Condensé, CEA-Saclay, F-91191 Gif-sur-Yvette, France

Single-wall carbon nanotubes^{1,2} are ideally suited for electron-transport experiments on single molecules because they have a very robust atomic and electronic structure and are sufficiently long to allow electrical connections to lithographically defined metallic electrodes. The electrical transport properties of single nanotubes³ and bundles of nanotubes⁴ have so far been interpreted by assuming that individual electrons within the nanotube do not interact, an approximation that is often well justified for artificial mesoscopic devices such as semiconductor quantum dots⁵. Here we present transport spectroscopy data on an individual carbon nanotube that cannot be explained by using independent-particle models and simple shell-filling schemes. For example, electrons entering the nanotube in a low magnetic

field are observed to all have the same spin direction, indicating spin polarization of the nanotube. Furthermore, even when the number of electrons on the nanotube is fixed, we find that variation of an applied gate voltage can significantly change the electronic spectrum of the nanotube and can induce spin flips. The experimental observations point to significant electron–electron correlations. We explain our results phenomenologically using a model that assumes that the capacitance of the nanotube depends on its many-body quantum state.

Figure 1a shows an atomic force microscope (AFM) image of an individual metallic carbon nanotube, with a measured height of 1.4 nm, contacting two Al/Pt electrodes that are separated by 200 nm. The sample was fabricated as described elsewhere^{3,6}. At room temperature the total resistance of the circuit is ~ 2 M Ω , with a contact resistance of the order of 500 k Ω . All presented measurements were performed in a dilution refrigerator with a mixing chamber base temperature of 5 mK. We apply a bias voltage (V_{bias}) to these electrodes and measure the d.c. current (I) through the nanotube. A third electrode at a distance of ~ 3 μm from the tube (outside the image) is used as the gate electrode to vary the electrostatic potential of the tube. Detailed transport spectroscopy measurements could be performed on this sample owing to its improved stability and reproducibility compared to our previous samples.

In Fig. 1b the differential conductance dI/dV_{bias} is displayed in colour as a function of V_{bias} and gate voltage (V_{gate}). The data reproduce many of the features found in previous electron transport experiments on metallic single-wall carbon nanotubes^{3,4}. For example, when V_{gate} is swept at low V_{bias} , an almost periodic sequence of

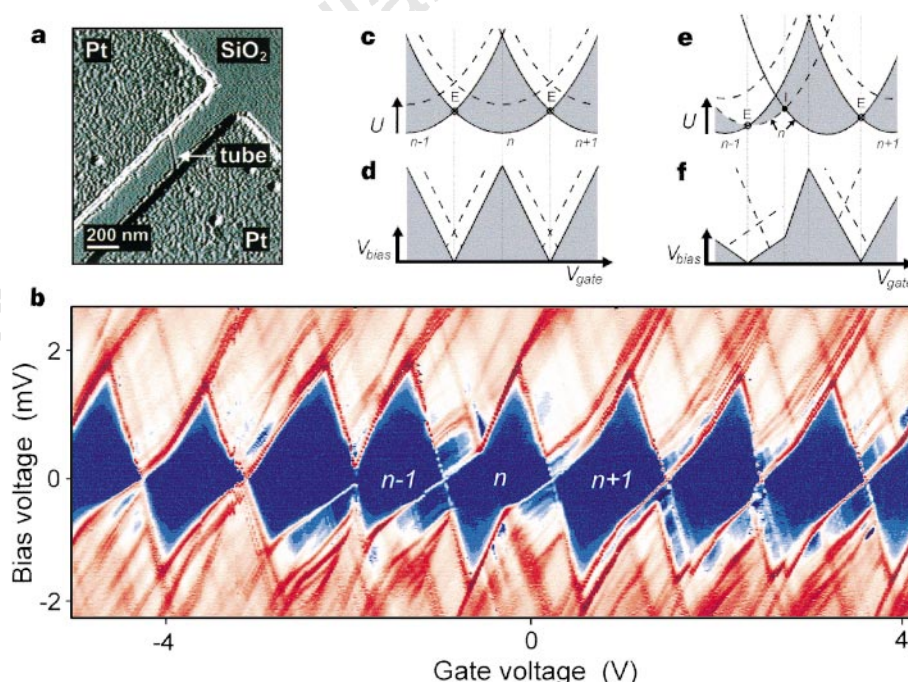


Figure 1 Transport spectrum of a single carbon nanotube at low temperature. **a**, AFM tapping-mode image of an individual carbon nanotube of 1.4 nm height, on top of a Si/SiO₂ substrate with electrodes of 25-nm-thick Al, capped with 5-nm Pt. A gate-electrode is positioned at a distance of 3 μm , outside the image. During the measurements, a minimum magnetic field B of 200 mT was applied perpendicular to the substrate in order to drive the Al from superconducting to normal. **b**, Differential conductance dI/dV_{bias} as a function of V_{bias} and V_{gate} : blue represents $dI/dV_{\text{bias}} = 0$, while red corresponds to a high dI/dV_{bias} . The data have been obtained by measuring many $I - V_{\text{bias}}$ curves for different V_{gate} . **c**, Energy diagram for a system with independent electrons. The energy U , equation (1), is plotted as a function of V_{gate} for ground states (solid lines) with $n - 1$, n , $n + 1$ excess electrons on the tube. Dashed parabolas represent excited states. Current can flow through the circuit when electrons can jump on and off the

tube, that is, when transitions occur between parabolas of consecutive electron number (for example, n and $n + 1$). At low V_{bias} such an 'external' transition occurs at the points E. In between the two points E, the minimum V_{bias} for an external transition is proportional to the difference in energy between the two respective parabolas (grey area). **d**, Transition diagram in terms of V_{gate} and V_{bias} corresponding to **c**. The solid lines separate the coulomb blockade (CB) region where no current flows (grey) and the conduction regions where a current flows. Their slope is determined by the horizontal distance between the parabolas, which are equal here. **e**, Energy diagram when the excited-state parabola for n electrons is displaced with respect to the ground-state parabola. Now there is a discontinuity in the ground-state energy at point I, which marks an 'internal' transition between two states. **f**, Corresponding transition diagram which yields features in the CB region that are qualitatively similar to the experimental data of **b**.

sharp conductance peaks occurs, which is a signature of Coulomb blockade (CB) of the tube⁷. In the blue regions between these Coulomb peaks, no current can flow through the device because adding just a single electron will cost too much energy. In such a CB region, the system remains in its ground state with a fixed number n of excess electrons on the tube. At a Coulomb peak, transitions occur between ground-states of different electron number. At higher V_{bias} beyond the CB regions, additional lines are observed, signalling a stepwise increase of the current. The current steps result from transitions involving excited states that are well separated from the ground state. The energy spectrum of the tube thus is not continuous but discrete, as is expected for a tube of finite length.

We model the measured spectrum by considering the energy U of the system as a function of the external voltages V_{gate} and V_{bias} and the number of electrons n_r and n_l which have flowed through the right and the left junction, respectively. Assuming equal junction capacitances and symmetric voltage bias, the terms of the energy which are relevant here for a state with $n = n_r - n_l$ excess electrons on the tube are:

$$U = E_c \left(n - \frac{C_g V_{\text{gate}}}{e} \right)^2 - \frac{n_r + n_l}{2} V_{\text{bias}} + E_{\text{kin}} \quad (1)$$

where $E_c = e^2/2C_\Sigma$ is the Coulomb energy for one additional electron. The symbols C_Σ , C_g denote respectively the total capacitance to the tube and the gate capacitance, and E_{kin} is the kinetic electronic energy of the state. As a function of V_{gate} , the energy U is a parabola. In Fig. 1c we schematically draw in solid lines the parabolas of three ground states with electron numbers $n-1$, n and $n+1$, which are equidistant. This diagram resembles a 'reaction coordinate diagram', which is often used to describe chemical reactions. In our case, the reaction coordinate V_{gate} is a voltage which is fixed externally. From this energy diagram one can deduce the conditions for current flow through the device by tracing possible transitions between parabolas of different electron number. The solid lines in Fig. 1d give these conditions for transitions, where electrons can jump on and off the nanotube, in terms of V_{bias} and V_{gate} . The grey area denotes the CB region, while the dashed lines represent excited-state transitions. Measurements on mesoscopic devices, such as semiconductor quantum dots and single-electron transistors, indeed yield such triangle-shaped CB regions⁵.

When we compare the data of Fig. 1b with this model, striking deviations are observed. The transition lines in the data do not have identical slopes, even for a single CB region. Kinks and bends in the lines are frequent. The kinks in the CB region labelled n are particularly clear. Such a kink indicates that the system makes a transition to a different ground state. As the excess charge n at this transition does not change, we call it an 'internal' transition. For 'external' transitions, n does vary and current flows through the tube.

We propose a new model to explain these data qualitatively. We make the hypothesis that C_g is not solely dependent on the geometric shape of the nanotube and electrodes but can be a function of the many-body state adopted by the tube. This complication is unnecessary for larger artificial metallic structures⁷ because external electric fields are perfectly screened inside the metal, leading to a state-independent capacitance. This last property is less and less verified as the dimensions of the structure become comparable to the screening length⁸. One may imagine that the electron densities of the one-particle wavefunctions have different spatial profiles. The various configurations of those states will therefore have varying capacitances. Another possibility is that electron-electron correlations make the effectiveness of screening dependent on the many-body state adopted by the tube. In our model, C_g thus depends both on n and on the degree of excitation of the tube, which induces a possible horizontal shift of the excited-state parabolas with respect to the corresponding ground-state parabolas. This is illustrated in Fig. 1e, which is similar to Fig. 1c, except that the two parabolas with the same electron number n are

shifted with respect to each other. It leads to an internal transition at point I where a sharp discontinuity in the ground-state energy occurs. The corresponding transition diagram (Fig. 1f) now appears to have transition lines with different slopes and a kink at the discontinuity, which qualitatively resembles the data. The more complex multiple kinks or bends in other parts of Fig. 1b indicate that the complete spectrum is more complicated than in the diagram of Fig. 1f. More than two parabolas could be involved, and the position of the parabolas may be dependent on V_{bias} , which is not taken into account in this model.

In Fig. 2a we follow the position of the Coulomb peaks along the V_{gate} -axis as a function of magnetic field B between 0 T and 8 T. As the field is increased, the parabolas move up or down in energy depending on their spin due to the Zeeman effect, leading to shifts in the position of the Coulomb peaks (points E in Fig. 1e) along the V_{gate} axis. It has been shown that, in nanotubes, the Zeeman effect dominates the magnetic-field dependence and that orbital effects may be neglected^{3,9}. This means that we can directly measure the spin direction of the incoming electron by the sign of the slope. From the average width of the Coulomb peaks at $V_{\text{bias}} = 100 \mu\text{V}$ (data not shown), we can translate shifts along the V_{gate} axis to energy variations and obtain an average g -factor of 2 ± 0.5 . The slopes dV_{gate}/dB of the different trace-sections in Fig. 3a appear to vary up to 25%. Such behaviour is expected in a model of shifted

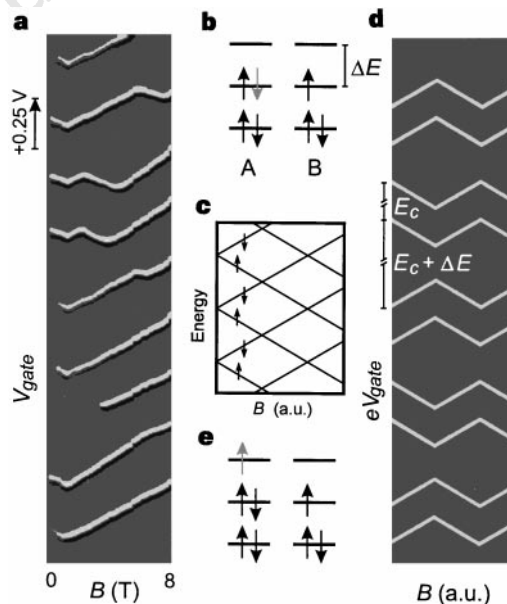


Figure 2 Evolution of the Coulomb peak positions as a function of the applied magnetic field, demonstrating spin-polarized states in the nanotubes. **a**, Numerical derivative dI/dV_{gate} is plotted in grey-scale as a function of V_{gate} and B , for $V_{\text{bias}} = 20 \mu\text{V}$. The top trace is counted as the first trace and the bottom trace as the ninth trace. For clarity, all traces have been moved towards each other by an equal amount. The seventh peak has a very low intensity at low magnetic field, but is here seen to develop in a peak with normal intensity after 4 T. The shifts of the Coulomb peak positions are a result of the Zeeman effect. **b**, Two ladders of degenerate states (A and B) of the tube, with level separation ΔE . For an independent-particle filling scheme, similar configurations are found every four added electrons. After adding two 'up' spins, the next incoming electron (grey) must have a 'down' spin. **c**, Magnetic field evolution of the doubly degenerate levels corresponding to the filling indicated in **b**. **d**, Expectation for measurements corresponding to **c**: the traces would have an additional separation corresponding to the Coulomb energy E_c for each extra electron on the tube. The data show a different pattern of traces (**a**). From 0 to 1.3 T, the energy of all the visible transitions goes down in energy, indicating that the tube adopts non-trivial spin-polarized states. **e**, Possible filling scheme suggested by the data (**a**). The incoming electron (grey) that has spin 'up' must occupy a level with a higher kinetic energy.

parabolas (Fig. 1e, f) because the speed of the horizontal shift of point E now also depends on the horizontal distance between the parabolas.

The data indicate non-trivial shell-filling and spin polarization in carbon nanotubes. Nanotubes have a simple shell structure^{10–12} with just two one-dimensional modes A and B. They are represented in Fig. 2b as two degenerate ladders of orbitals. Considering regular shell filling of independent particles, one would expect an up–down spin filling yielding a simple evolution of the states (Fig. 2c) and a corresponding movement of Coulomb-peak positions (Fig. 2d). In the data, we observe striking differences from this expected pattern. First, all the eight visible peaks appear to go down in energy from 0 T to ~1.3 T, which indicates that all electrons enter the tube with their spin parallel to the field. After 1.3 T, the peaks predominantly go up in energy. Second, adjacent traces do not show the expected correlation as indicated in Fig. 2d. Although paired lines are observed (for example, lines 3 and 4; see Fig. 2 legend for nomenclature), traces adjacent to this pair do not have kinks in opposite directions at the same magnetic field, which is a characteristic feature in quantum dots^{5,13}. This implies that, at a given magnetic field, the $(n + 1)$ -electron ground state is not simply the n -electron ground state with one electron added in a new orbital.

In Fig. 3 we closely examine the second trace of Fig. 2a by plotting the numerical derivative dI/dV_{gate} of this peak at $V_{\text{bias}} = 500 \mu\text{V}$ as a

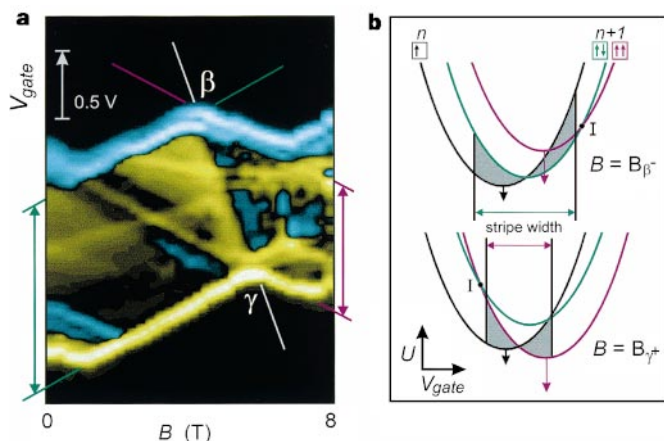


Figure 3 Evolution of the excited-state spectrum with magnetic field. **a**, Numerical derivative dI/dV_{gate} of trace 2 of Fig. 2a as a function of V_{gate} and B for $V_{\text{bias}} = 500 \mu\text{V}$. Yellow lines correspond to transitions entering the bias window, while blue lines correspond to transitions leaving the window. The vertical widths in V_{gate} of the second (green) and the third (purple) segment are unequal. To model this unusual result qualitatively, we drawn in **b** a possible configuration of shifted parabolas just before kink β ($B = B_{\beta^-}$) and just after kink γ ($B = B_{\gamma^+}$). In **b** (top), the width of the second section of the stripe (green lines in **a**) is given by the horizontal width of the grey area, where transitions between the n -electron ground state with $m_s = 1/2$ (black parabola) and the $(n + 1)$ -electron ground state with $m_s = 0$ (green) can occur. The horizontal width of the grey area is determined by eV_{bias} , which is the maximum vertical distance between parabolas where transitions between the parabolas can occur. In going from β to γ , the parabolas shift with a speed indicated by the arrows. The triplet state moves twice as fast as the doublet, while the singlet state is unchanged. In **b** (bottom), the stripe is now narrower (purple lines in **a** and grey area for $B = B_{\gamma^+}$) because the $n + 1$ ground state is now the purple parabola with $m_s = 1$, which is further away in V_{gate} from the black parabola than the green parabola. We note that the internal transition point I between the $n + 1$ parabolas is shifted through the bias window from just to the right of the grey current area to just to the left with B . The line connecting β and γ in **a** traces this point I. In the traditional model (Fig. 1c), parabolas with the same n are positioned on top of each other, which yields a strictly vertical β – γ line. A non-vertical β – γ line signals a shift between parabolas associated with different spins.

function of B . Because of the higher bias voltage, the trace has broadened into a stripe with additional structure due to excited-state transitions. Along the B -axis, the stripe consists of four segments, of which the second and the third appear to have clearly different widths in V_{gate} . This is very unusual for such measurements, and will be discussed below. As Fig. 2c shows, one expects levels to split from 0 T, which should also be observable in the excited-state spectrum as a line with a positive slope intersecting the lowest yellow line at 0 T in Fig. 3a. Such a feature is not observed in this Coulomb peak or in others. This lack of spin-splitting at 0 T reproduces our earlier results on an individual nanotube³. In contrast, transport studies in a bundle of nanotubes⁹ have recently yielded the expected Zeeman splitting at 0 T. Our data also indicate that the patterns of excited-state transitions versus B (for example, Fig. 3a) are not correlated with the patterns of ground-state transitions (Fig. 2a), which is unexpected in a model of independent particle states.

We argue that the changing width of the stripe at the β – γ kink (Fig. 3a) is the consequence of an internal transition similar to the ones signalled by the kinks in the $V_{\text{bias}} - V_{\text{gate}}$ plot (Fig. 1b). The β – γ kink results from a crossing of two levels with different spins (see Fig. 3b). As the line connecting the two kinks β and γ is not vertical (which would lead to equal stripe widths), we know that the two corresponding parabolas are shifted with respect to each other. This in turn implies that at the crossing of the parabolas, an internal transition of a magnetic nature occurs. A curious situation thus occurs: changing the gate voltage results in a spin flip in the tube when the β – γ line is crossed. We suppose that such gate-voltage-driven internal magnetic transitions also exist down to $B = 0$. For example, we may extrapolate internal transition lines like the β – γ line to 0 T. In such a scenario, we can explain the fact that all the traces in Fig. 2a initially all go down with increasing magnetic field. Upon increasing gate voltage, we envisage a sequence for the total spin of the tube such as:

$$0 \rightarrow \frac{1}{2} \rightarrow 1 \triangleright 0 \rightarrow \frac{1}{2} \rightarrow 1 \triangleright 0 \rightarrow \dots$$

where $\rightarrow$ marks an external transition increasing the number n of electrons by 1, and $\triangleright$ a magnetic internal transition where n stays constant. We note that at an external transition the spin is always increased, but between Coulomb peaks a hidden internal magnetic transition occurs which lowers the total spin of the tube. The spins of all incoming electrons can thus be parallel without having to invoke the existence of a gigantic spin polarization. This hypothesis implies that for certain gate voltage ranges, the ground state of the tube has a spin angular momentum higher than $1/2$. In the simple sequence described above, the highest spin state would be a triplet state. Such a triplet state is shown in Fig. 2e where an incoming electron (grey) occupies an orbital with a higher kinetic energy. We now again consider the origin of the dependence of the capacitance on the quantum state of the nanotube and the resulting parabola displacement. We believe that a non-perfect and state-dependent screening due to electron–electron interactions is the cause of the parabola displacement, because models based on different spatial profiles of single-particle states do not produce the observed spin flips.

Thus, we have observed five features in the magnetic-field dependence of Coulomb peaks that cannot be explained by a simple shell-filling scheme of one-electron states. (1) At low fields, electrons always enter the nanotube with the same spin. (2) Kinks in adjacent traces of Coulomb peak position versus field do not occur at the same field. (3) The evolution of excited-state transitions with B is uncorrelated with the evolution of ground-state transitions with field. (4) Zeeman splitting at 0 T is not observed. (5) The gate voltage can induce spin flips. The first observation can be explained in a framework of one-electron states only if a very large exchange-type electron–electron correlation is present. For observations (2) to (5), however, a model based on one-electron states

does not suffice. These observations indicate that many-body states due to significant electron–electron correlations beyond exchange-type correlations have to be taken into account to describe the low-energy transport properties of carbon nanotubes. □

Received 5 February; accepted 11 June 1998.

- Bethune, D. S. *et al.* Cobalt-catalysed growth of carbon nanotubes with single-atomic-layer walls. *Nature* **363**, 605–607 (1993).
- Iijima, S. & Ishihashi, T. Single-shell carbon nanotubes of 1-nm diameter. *Nature* **363**, 603–605 (1993).
- Tans, S. J. *et al.* Individual single-wall carbon nanotubes as quantum wires. *Nature* **386**, 474–477 (1997).
- Bockrath, M. *et al.* Single-electron transport in ropes of carbon nanotubes. *Science* **275**, 1922–1925 (1997).
- Sohn, L. L., Kouwenhoven, L. P. & Schön, G. (eds) *Mesoscopic Electron Transport* (Kluwer Academic, Dordrecht, 1997).
- Thess, A. *et al.* Crystalline ropes of metallic carbon nanotubes. *Science* **273**, 483–487 (1996).
- Grabert, H. & Devoret, M. H. (eds) *Single Charge Tunneling* (Plenum, New York, 1992).
- Hallam, L. D., Weiss, J. & Maksym, P. A. Screening of the electron–electron interaction by gate electrodes in semiconductor quantum dots. *Phys. Rev. B* **53**, 1452–1462 (1996).
- Cobden, D. H. *et al.* Spin splitting and even–odd effects in carbon nanotubes. *Phys. Rev. Lett.* **81**, 681–684 (1998).
- Mintmire, J. W., Dunlap, B. I. & White, C. T. Are fullerene tubules metallic? *Phys. Rev. Lett.* **68**, 631–634 (1992).
- Hamada, N., Sawada, A. & Oshiyama, A. New one-dimensional conductors: graphitic microtubules. *Phys. Rev. Lett.* **68**, 1579–1581 (1992).
- Saito, R., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. Electronic structure of chiral graphene tubules. *Appl. Phys. Lett.* **60**, 2204–2206 (1992).
- Tarucha, S., Austing, D. G., Honda, T., van der Hage, R. J. & Kouwenhoven, L. P. Shell filling and spin effects in a few electron quantum dot. *Phys. Rev. Lett.* **77**, 3613–3616 (1996).

Acknowledgements. We thank R. E. Smalley and co-workers for the supply of the indispensable single-walled carbon nanotubes. We also thank L. P. Kouwenhoven, T. H. Oosterkamp and Yu. V. Nazarov for discussions, and B. van den Enden for technical assistance. This work was supported by the Dutch Foundation for Fundamental Research on Matter (FOM).

Correspondence and requests for materials should be addressed to C.D. (e-mail: dekker@qt.tn.tudelft.nl).

Pairwise selection of guests in a cylindrical molecular capsule of nanometre dimensions

Thomas Heinz, Dmitry M. Rudkevich & Julius Rebek Jr

The Skaggs Institute for Chemical Biology and the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

'Container' complexes in which a guest molecule is held mechanically within a cage-like host have been known for over a decade^{1,2}. They provide a means to stabilize reactive intermediates³ and to create new forms of stereoisomerism⁴. Molecular capsules held together by hydrogen bonds are more recent; they are formed reversibly on timescales of milliseconds to hours, long enough for molecular motions⁵ and even reactions⁶ to be seen for the encapsulated species. Here we describe the synthesis and characterization of a hydrogen-bonded molecular capsule of nanometre dimensions, which is large enough to encapsulate two different molecules. This allows us to explore the size- and shape-selectivity of the encapsulation process: we see, for example, the exclusive formation of the hetero-guest pair when benzene and *p*-xylene are both added to a solution. This presumably reflects optimal occupancy of the capsule—two benzene guests leave too much empty space in the interior, and two *p*-xylene molecules make it too crowded.

The synthesis of the new construct **1** (Fig. 1) uses methods developed by Cram⁷ for the synthesis of velcra (compounds whose molecules stick to one another) but leads to molecules of starkly different shapes and behaviours. Whereas the velcra form dimeric structures through face-to-face aryl stacking interactions⁷, compound **1** dimerizes through hydrogen bonding at its edges. The shape of **1** is vase-like, and the dimerization takes place in the rim-to-rim manner to give the large cylindrical capsule **1**·**1** (Fig. 1). It features dimensions $\sim 1.0 \times 1.8$ nm and is

stabilized by a 'seam' of eight bifurcated hydrogen bonds, with the imide hydrogens of one molecule directed between two carbonyl oxygens of another. This cyclic pattern of imides is, to our knowledge, without precedent.

The ¹H NMR spectra of **1**·**1** in CDCl₃, benzene-*d*₆ or toluene-*d*₈, show one sharp set of signals for all groups of protons (Fig. 2a, b and Fig. 3a), characteristic of C_{4v} symmetry of the subunits^{8,9}. The spectra do not change within a 220–330 K temperature interval (toluene-*d*₈, CDCl₃), indicating high conformational stability. The NMR spectra in these solvents feature N-H signals far downfield at >9.5 p.p.m. and the Fourier-transform infrared spectrum of **1** in CHCl₃ shows only hydrogen-bonded N-H stretching absorption as a broad band at 3,332 cm⁻¹ at > 0.3 mM concentrations.

Molecular modelling¹⁰ suggested that two benzene or two toluene molecules could easily be accommodated inside the capsule, but the larger solvent mesitylene-*d*₁₂ did not seem to fit well within the capsule. Some of the hydrogen bonds of the capsule become disrupted by the solvent, and indeed, multiple N-H signals were observed in the downfield region of the spectra in this solvent (Fig. 2c). This augured well for encapsulation studies in mesitylene-*d*₁₂, and the addition of suitable guests, such as bibenzyl, terphenyl, dicyclohexyl carbodiimide and *trans*-stilbene in that solvent restored the high symmetry of the capsule as reported by the NMR spectra (see for example, Fig. 4a). In addition, a second set of signals for the guest appeared, and their upfield shifts are characteristic of species encapsulated in aromatic containers¹. The integration of the ¹H NMR spectra of these signals established the 1:1 stoichiometry of the host–guest assembly.

The encapsulation of less symmetrical guests such as *N*-phenyl benzylamine, benzyl phenyl ether, *trans*-4-stilbene methanol, and *p*-[*N*-(*p*-tolyl)]toluamide (Fig. 4b, c) shows the effect of molecular length on the rotational freedom of the guest and the overall symmetry of the complex. The NMR spectrum of the complexes shows a doubling of the capsule's signals: the two ends are different (Fig. 4b, c). These long guests can spin inside along the axis of the capsule, but are too large to 'tumble' within it. This restriction of motion is featured in the stereoisomerism discovered by Reinhoudt⁴ in covalently bound carceplexes, and was also observed by Sherman¹¹ in reversibly formed capsules.

The capacity of **1**·**1** for small molecules—even common solvents—could also be directly observed in competition with mesitylene-*d*₁₂. In this solvent, the addition of (non-deuterated) toluene gave rise to a single set of sharp signals for the capsule in the NMR spectrum. Integration showed that two toluenes occupied the capsule (Fig. 4d). Toluene is small enough to tumble rapidly within the capsule on the NMR timescale at room temperature, and an averaged signal is observed for the methyl groups at –0.78 p.p.m.

The selectivity of the capsule was revealed through competition experiments involving two solvent guests. When both benzene and *p*-xylene were added in a 1:1 ratio (~30-fold excess) to the mesitylene-*d*₁₂ solution of **1**, an asymmetrically filled capsule (see also Fig. 3b) was observed exclusively. One benzene (a singlet at 3.67 p.p.m.) and one xylene molecule (two aromatic doublets at 5.48 and 3.08 p.p.m.; *J* = 6.3 Hz, and one of two methyl singlets at –2.81 p.p.m.; another one is hidden) were found encapsulated. Apparently, the capsule with two xylene molecules is too crowded; both modelling¹⁰ and ¹H NMR spectroscopy suggested that two xylenes cannot easily fit inside. Instead, a comfortable occupancy is reached with one of each guest in the capsule (Fig. 3b). The complex is asymmetrical because the two guests cannot squeeze past each other to exchange positions in the capsule—at least not on the NMR timescale. Likewise, benzene paired with *p*-trifluoromethyltoluene, *p*-chlorotoluene, 2,5-lutidine and *p*-methylbenzyl alcohol (Fig. 3c) to give species with one of each guest inside. The chemical shift of

the guest methyl groups (-2.7 to -2.8 p.p.m.) places them near the ends of the capsule.

We performed a competition experiment involving three different solvent guests, in which toluene, benzene and *p*-xylene were

added in a 2:1:1 ratio (~ 60 -fold excess of toluene, ~ 30 -fold excess of benzene and *p*-xylene) to the mesitylene- d_{12} solution of **1**. Again, the capsule was filled preferentially ($\sim 90\%$) with benzene and *p*-xylene, and only $\sim 10\%$ of the capsule with two toluenes inside

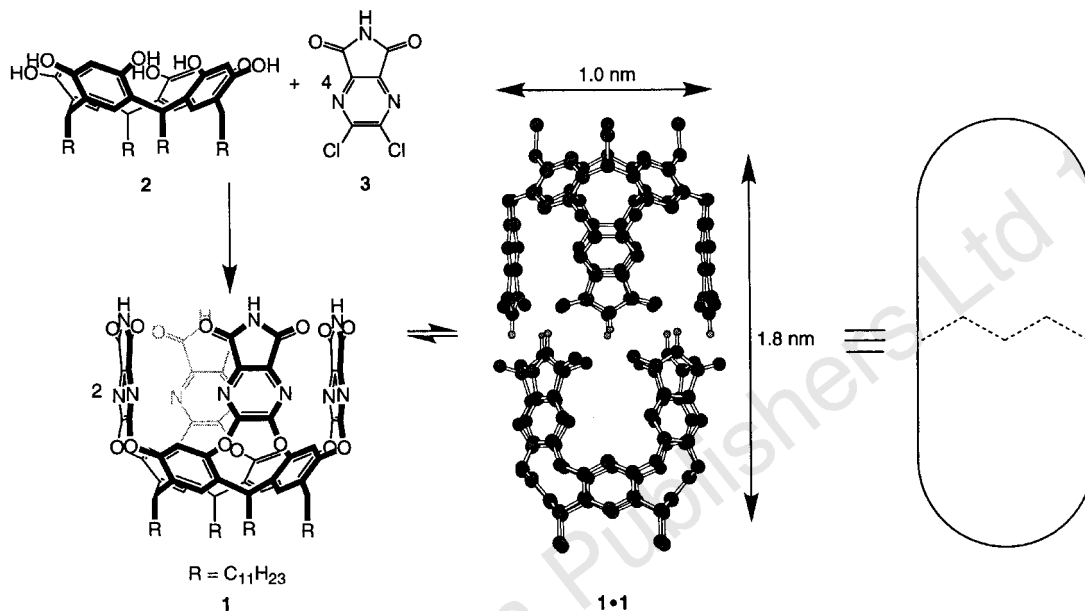


Figure 1 Synthesis and structure of cavitand **1**. The synthesis of cavitand **1** involves coupling of *C*-undecylcalix[4]resorcinarene **2** with 5,6-dichloropyrazine-2,3-dicarboxylic acid imide **3** in dimethylformamide (DMF) in the presence of NEt_3 (48% yield). Two cavitands self-assemble through hydrogen bonding into a

cylindrical capsule **1·1**. Two representations of the capsule are shown. In the centre is the energy-minimized¹⁰ (MacroModel 5.5, Amber^{*} force field) structure **1·1**; the long alkyl chains and CH hydrogen atoms are omitted for viewing clarity. Right, the cartoon representation used in Figs 3 and 4.

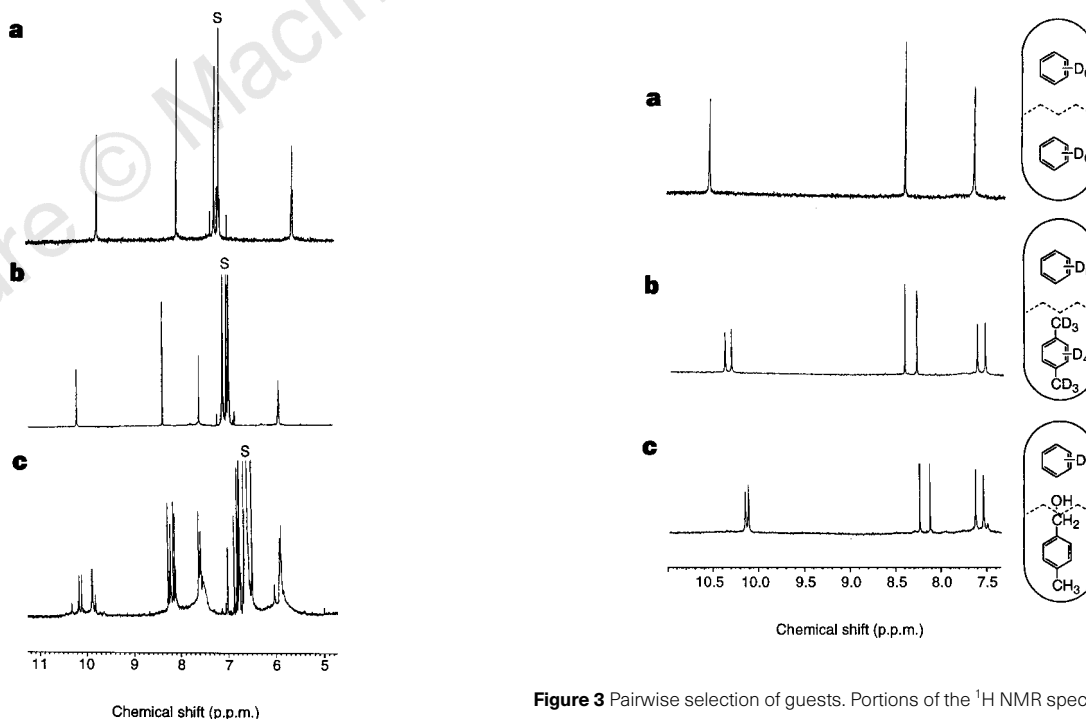


Figure 2 ^1H NMR spectra of compound **1** in different solvents. Downfield portions of the ^1H NMR spectra (600 MHz, 295 K) of compound **1** (concentration ~ 1 mM). **a**, in CDCl_3 , δ 9.79 (s, 4 H, NH), 8.09 (s, 4 H, arom), 7.37 (s, 4 H, arom), 5.64 (t, 4 H, $J = 7.9$ Hz, CH-methine). Other signals (not shown): δ 2.25 (m, 8 H, alkyl), 1.20 (m, 72 H, alkyl), 0.83 (t, 12 H, $J = 7.5$ Hz, CH_3). **b**, In toluene- d_8 ; and **c**, in mesitylene- d_{12} . Characteristically for the C_{4v} conformation^{8,9}, the methine CH triplet is situated at ~ 6 p.p.m. The residual solvent signals are marked S. The ^1H NMR spectrum in dimethylsulphoxide- d_6 shows significant broadening of the calix[4]resorcinarene skeleton at 295 K. The sharp spectrum was obtained at >350 K.

Figure 3 Pairwise selection of guests. Portions of the ^1H NMR spectra (600 MHz, 295 K) of capsule **1·1** (concentration of **1** ~ 1 mM). **a**, in benzene- d_6 , δ 10.50 (s, 4 H, NH), 8.37 (s, 4 H, arom), 7.60 (s, 4 H, arom). Other signals (not shown): δ 5.92 (t, 4 H, $J = 7.2$ Hz, CH-methine), 2.30 (m, 8 H, alkyl), 1.34 (m, 72 H, alkyl), 0.98 (t, 12 H, $J = 7.1$ Hz, CH_3); **b**, in benzene- d_6 -*p*-xylene- d_{10} , 1:1. When benzene- d_6 and *p*-xylene- d_{10} were employed at different molar ratios, namely 1:70, 1:3, 1:1, and 3:1, respectively, only the asymmetrically filled capsule was observed. However, when the ratio benzene- d_6 -*p*-xylene- d_{10} of 70:1 was employed, both the benzene-filled (**a**) and the asymmetrically filled (**b**) capsule were present. **c**, (1-benzene- d_6 -*p*-methylbenzyl alcohol-**1**) in mesitylene- d_{12} . The CH_3 singlet of the complexed *p*-methylbenzyl alcohol is situated at -2.77 p.p.m.



Figure 4 Guest encapsulation experiments. Portions of the ^1H NMR spectra (600 MHz, 295 K) of complexes (**1**·**1**-guest) in mesitylene- d_{12} (concentration of **1** ~1 mM, guest concentration >50 mM): **a**, (**1**·**1**-dicyclohexyl carbodiimide). Most of the well-resolved cyclohexyl signals for the encapsulated guest are situated between 0 and -4 p.p.m.; **b**, (**1**·**1**-*trans*-4-stilbene methanol). The CH_2OH fragment of the encapsulated guest molecule is situated at -0.24 p.p.m. (d, 2 H, $J = 3.5$ Hz, CH_2) and -3.33 p.p.m. (t, 1 H, $J = 3.5$ Hz, OH). After the addition of D_2O , the imide

NH singlet of **1** and the OH triplet of the encapsulated guest exchanged within a few hours; **c**, **1**·**1**-*p*-[*N*-(*p*-tolyl)]toluamide. The CH_3 groups of the encapsulated guest are found at -2.70 and -2.81 p.p.m.; **d** (**1**·**1**·2 × toluene). The CH_3 singlet of the two complexed toluenes at -0.78 p.p.m. broadened and disappeared into the baseline at temperatures below 273 K. The observed spectrum indicates a fast tumbling of the toluenes within the capsule on the NMR timescale, leading to an averaged signal for the three diastereomers shown.

was observed. In another experiment, benzene and *p*-xylene in mesitylene- d_{12} replaced encapsulated toluenes within a few minutes at room temperature (using a 1:1:1 ratio of toluene, benzene and *p*-xylene).

It was not expected that molecule-within-molecule complexes, held together only by hydrogen bonds, would show such selectivity, and matching the overall length of two guest molecules with the dimensions of the cavity appears to be a behaviour unique to these capsules. But taken together, these results point directly to the use of capsules as selective reaction chambers for bimolecular processes, and suggest that orientations of the reacting partners could be controlled in the future. □

Received 6 April; accepted 17 June 1998.

1. Cram, D. J. & Cram, J. M. *Container Molecules and Their Guests* (RSC, Cambridge, MA, 1994).
2. Collet, A. *et al.* Cyclotrimeric aryls and cryptophanes: their synthesis and applications to host-guest chemistry and to the design of new materials. *Top. Curr. Chem.* **165**, 103-129 (1993).
3. Warmuth, R. *o*-Benzynes: strained alkyne or cumulene?—NMR characterization in a molecular

container. *Angew. Chem. Int. Edn Engl.* **36**, 1347-1350 (1997).

4. Timmerman, P., Verboom, W., van Veggel, F. C. J. M., van Duynhoven, J. P. M. & Reinhoudt, D. N. A novel type of stereoisomerism in calix[4]arene-based carceplexes. *Angew. Chem. Int. Edn Engl.* **33**, 2345-2348 (1994).
5. Rebek, J. Jr Assembly and encapsulation with self-complementary molecules. *Chem. Soc. Rev.* 255-264 (1996).
6. Kang, J. & Rebek, J. Jr Acceleration of a Diels-Alder reaction by a self-assembled molecular capsule. *Nature* **385**, 50-52 (1997).
7. Cram, D. J., Choi, H.-J., Bryant, J. A. & Knobler, C. B. Solvophobic and entropic driving forces for forming velcroplexes, which are four-fold, lock-key dimers in organic media. *J. Am. Chem. Soc.* **114**, 7748-7765 (1992).
8. Moran, J. R. *et al.* Vases and kites as cavitands. *J. Am. Chem. Soc.* **113**, 5707-5714 (1991).
9. Rudkevich, D. M., Hilmersson, G. & Rebek, J. Jr Intramolecular hydrogen bonding controls the exchange rates of guests in a cavitand. *J. Am. Chem. Soc.* **119**, 9911-9912 (1997).
10. Mohamadi, F. *et al.* MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **11**, 440-467 (1990).
11. Chapman, R. G. & Sherman, J. C. Study of templation and molecular encapsulation using highly stable and guest-selective self-assembling structures. *J. Am. Chem. Soc.* **117**, 9081-9082 (1995).

Acknowledgements. We thank the Skaggs Research Foundation and the National Institutes of Health for financial support. T.H. is grateful to the Swiss National Science Foundation and the Ciba-Geigy-Jubiläums-Stiftung for the fellowships.

Correspondence and requests for materials should be addressed to J.R. (e-mail: jrebek@scripps.edu).

A million-year record of fire in sub-Saharan Africa

M. I. Bird & J. A. Cali

Research School of Earth Sciences, Australian National University, Canberra, ACT 0200, Australia

Biomass burning today constitutes approximately one-third of annual anthropogenic CO₂ emissions, and there is a sound theoretical base for expecting fire-related changes in vegetation patterns to affect climate, at least on a regional scale^{1–3}. But despite the central role that fire has played in moulding many modern ecosystems, there is little information on the incidence of fire before the earliest time at which anthropogenic burning may have significantly affected natural fire regimes. Here we present a million-year record of elemental carbon abundance from marine sediments on the Sierra Leone rise, 'downwind' of sub-Saharan Africa. Elemental carbon serves as a proxy for wind-blow debris derived from the combustion of sub-Saharan vegetation. The inferred fire incidence in the region was low until about 400,000 years ago, but since that time intense episodes of vegetation fires have occurred during periods when global climate was changing from interglacial to glacial mode. The occurrence of a peak in elemental carbon abundance within the present interglacial is unique in the past million years, suggesting that this peak is anthropogenic in origin, and that humans have exercised significant control over fire regimes in the region at least since Holocene times.

Combustion has underpinned the technological ascent of the genus *Homo*, and lies at the core of current concerns over global change, both through biomass burning, and combustion associated with the production of energy. The potency of fire as an agent of environmental change has led to speculation that prehistoric humanity may have been actively engaged in modifying vegetation patterns, regional climate and atmospheric trace-gas concentrations since long before the advent of modern concerns over global change⁴. For example, it has been suggested that prehistoric human use of fire, coupled with domestication of livestock, has led to the expansion of the Sahel in the past few thousands years, and an increase in recent desertification rates in the region^{5–7}. But such a view is not universally held, and others favour natural mechanisms^{8,9}.

Long records of fire frequency, integrated over a large continental area and spanning several glacial–interglacial cycles, can provide a baseline for natural variability and long-term trends in fire incidence. Deep-sea sediments are particularly suited to this kind of study because they can provide relatively undisturbed, continuous records of large-scale phenomena which are readily dated using oxygen-isotope stratigraphy.

Several chemical techniques have been proposed to quantify the abundance of elemental carbon in marine sediments^{10–14}, and these techniques have been used to isolate elemental carbon at the K/T boundary¹⁵, and to estimate its abundance in marine sediments of pre-industrial, Quaternary and Tertiary age^{11,16,17}.

We have modified methods developed by Wolbach and Anders¹² and Emiliani *et al.*¹³, and use two oxidative degradation steps to destroy organic carbon leaving a residue composed entirely of oxidation-resistant elemental carbon (OREC; see Fig. 1 legend). The OREC abundance in a sample is then determined manometrically after combustion to CO₂; this has the advantage that the carbon-isotope composition ($\delta^{13}\text{C}_{\text{OREC}}$ value) of the gas can be determined, providing information on the type of vegetation being burnt^{18,19}. Carbonate, aeolian dust, and amorphous silica contents, as well as organic carbon content and organic carbon $\delta^{13}\text{C}_{\text{OC}}$ value are also determined for each sample (Fig. 1).

Core ODP-668B from the Sierra Leone rise²⁰, in the equatorial eastern Atlantic Ocean, was chosen for analysis. The core-site is 'downwind' of sub-Saharan Africa, a region likely to have experienced large changes in fire incidence in response both to large-scale global climate changes and the more recent imposition of anthropogenic fire regimes. The products of modern biomass burning in sub-Saharan Africa have been detected as far west as Barbados, and a surface grab sample from the Sierra Leone rise has been reported to have 'clearly noticeable amounts of (elemental) carbon'²¹. Palaeoenvironmental reconstructions based on the aeolian dust content of marine sediments off sub-Saharan Africa suggest that the location

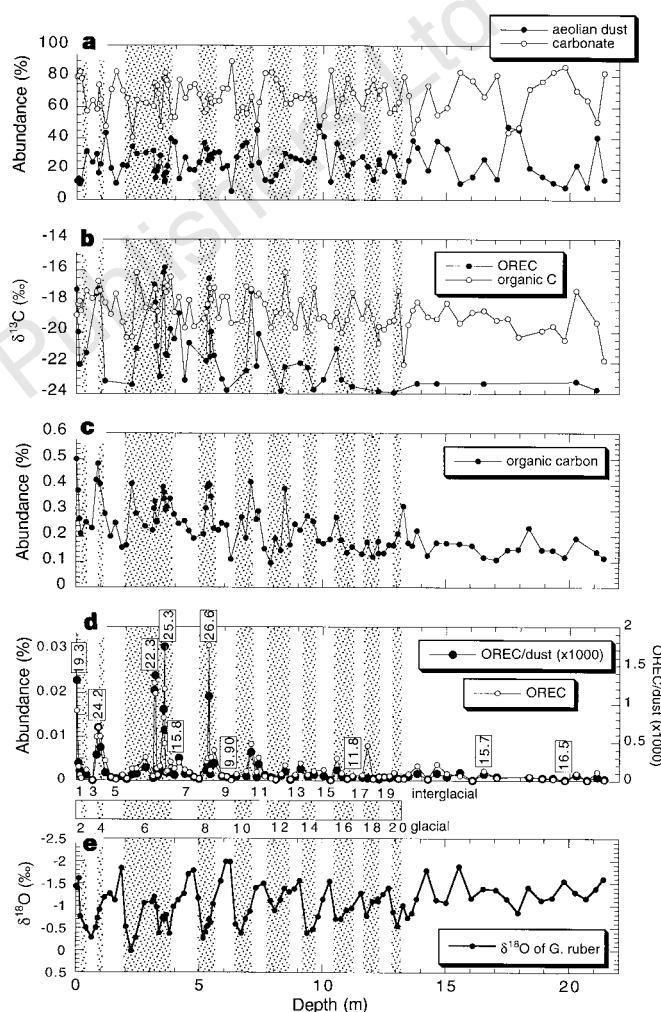


Figure 1 Sedimentological and isotopic results for ODP-668B, Sierra Leone rise. The site is located at 4° 46.23' N, 20° 55.62' W in 2,693 m of water²². The hole had a total length of 31.3 m with 100% recovery. Samples were prepared using the techniques discussed by Bird and Grocke²⁰. Shaded regions represent glacial intervals. **a**, Abundance of carbonate and aeolian dust as a percentage of dry weight. **b**, $\delta^{13}\text{C}$ value of OREC (oxidation-resistant elemental carbon) and bulk organic carbon $\delta^{13}\text{C}_{\text{OC}}$ ($\pm 0.1\text{‰}$). **c**, Abundance of bulk organic carbon as a percentage of dry weight ($\pm 2\%$). **d**, Abundance of OREC ($\pm 10\%$), given both as a percentage of dry weight, and as a ratio of the aeolian dust content of the same sample. Also shown (in boxes) are the C/N ratios of bulk organic matter for selected intervals. **e**, $\delta^{18}\text{O}$ value ($\pm 0.1\text{‰}$) of the planktic foraminifer *G. ruber* relative to VPDB. Oxygen-isotope stage assignments are based on comparison with a nearby high-resolution stratigraphy for Meteor 13519 (ref. 25). The stage assignments yield an approximately linear sedimentation rate for the entire sequence of 1.7 cm kyr⁻¹ compared with a slightly lower (1.4 cm kyr⁻¹), but also linear, rate of sediment accumulation in Meteor 13519. The Brunhes–Matuyama boundary²⁰ occurs at ~13.3 m depth in ODP-668B, also consistent with the stage assignments.

has been in the path of winds emanating from the sub-Saharan region since at least the last interglacial period^{22–24}. The location is on a topographic high, and sufficiently far from the coast to ensure that any signal is integrated over a large continental area and thus will not be sensitive to latitudinal shifts in vegetation type in response to climate changes.

The sediment is a muddy foraminiferal ooze containing 40–90% carbonate (Fig. 1a) and <5% amorphous silica (values not shown in Fig. 1), with the balance comprising aeolian dust. The upper 21 m of the core was sampled at an interval of 20 cm (5 cm over important intervals); representing over one million years of sediment accumulation. A study of a nearby core (Meteor-13519) by Sarinthein *et al.*²⁵ concluded that dust fluxes to the Sierra Leone rise have not varied greatly in the past.

Organic carbon contents in the core studied here are generally low, between 0.1% and 0.3% (Fig. 1c), gradually decreasing down the core as a result of diagenetic loss of labile carbon species²⁶. Several peaks in organic carbon abundance (up to 0.5%) occur in the upper half of the core, and similar peaks in other cores from the region have been related to increases in palaeoproductivity²⁷. $\delta^{13}\text{C}_{\text{OC}}$ values generally vary between -17‰ and -20‰ (Fig. 1b), and again such variations have been related to changes in productivity in the area in the past²⁶. The chronology for the core is based on the oxygen-isotope composition of the planktic foraminifer *Globigerinoides ruber* (Fig. 1e).

OREC abundances are expressed both as a raw percentage of dry weight, and as a ratio to the aeolian dust fraction. The second measure is the more informative, as it circumvents the potential problem of apparent changes in abundance being due to changes either in wind strength (particulate transport efficiency) or mass accumulation rate. Nevertheless, the general trends are the same using either measure (Fig. 1d) as OREC abundances vary by two orders of magnitude, far in excess of the variations in bulk sediment mass accumulation rate ($1.1\text{--}2.3\text{ g cm}^{-2}\text{ kyr}^{-1}$). Apparent OREC accumulation rates vary from $2\text{--}10\text{ }\mu\text{g cm}^{-1}\text{ kyr}^{-1}$ for periods when OREC deposition rates were low, up to $200\text{--}400\text{ }\mu\text{g cm}^{-1}\text{ kyr}^{-1}$ at times of high OREC deposition.

The abundances of OREC were generally low before the beginning of the stage 10 glacial. Following the small peak at the beginning of stage 10 ($\sim 400,000$ years before present) there are large peaks in abundance in stages 8, 6, 3–4 and 1. The early peaks are particularly sharp, occurring over intervals of 5–10 cm in the core; assuming a bioturbation depth of 5 cm (ref. 25), these peaks could represent short-duration intense events. The peak in stage 3–4 is lower, and occurs over a thicker sediment interval (~ 60 cm). Between the peaks, OREC abundances return to comparatively low levels. $\delta^{13}\text{C}_{\text{OREC}}$ values vary widely in the upper half of the core, from -25‰ when OREC abundances are low, to -15.8‰ when OREC abundances are high (Fig. 1b). This range of values is outside the range normally encountered for marine organic carbon in tropical regions, and well outside the range of $\delta^{13}\text{C}_{\text{OC}}$ values in the core.

Two additional lines of evidence suggest that the material being analysed is indeed elemental carbon, and not a resistant component of marine organic carbon.

First, the C/N ratios of organic carbon in samples from the intervals containing the OREC peaks are elevated above 20, while other samples from the core exhibit values as low as 9.8, similar to living marine organic matter (Fig. 1d). This suggests that a source high in carbon and low in nitrogen forms a significant component of the bulk organic carbon, consistent with that component being OREC.

Second, there is no relationship between the abundance or $\delta^{13}\text{C}$ value of OREC in a sample and either the percentage of organic carbon or the $\delta^{13}\text{C}_{\text{OC}}$ value of the same sample, as would be expected if the carbon being analysed was a resistant component of marine organic carbon. However, there is a strong relationship between the abundance of OREC and the $\delta^{13}\text{C}_{\text{OREC}}$ value (Fig. 2). This would be

expected if the carbon were indeed elemental carbon, as the incidence of fire in tropical regions is related to the distribution of savannah regions and hence to the distribution of C_4 grasses with high $\delta^{13}\text{C}$ values of $\sim -12\text{‰}$. Thus low fire incidence implies wet conditions and extensive C_3 -dominated forests ($\delta^{13}\text{C}$ values of $\sim -27\text{‰}$), and hence low $\delta^{13}\text{C}_{\text{OREC}}$ values. With the expansion of savannahs in drier conditions, fire incidence will increase, as will the proportion of C_4 grasses in the biomass being burnt, and hence $\delta^{13}\text{C}_{\text{OREC}}$ values will also increase.

The distribution of OREC peaks in the core is significant. The absence of significant peaks in OREC abundance before stage 10 suggests that either fire was not a common phenomenon in sub-Saharan Africa before this time, or that a shift in atmospheric circulation at this time first brought particulates emanating from the sub-Saharan region over the core location. Given that there are no large changes in the dust flux to the Sierra Leone rise²⁵, the former interpretation is preferred.

With the exception of the most recent peak, no OREC peaks occur either at glacial or interglacial maxima. Given that the sample interval was 20 cm, consideration should be given to the possibility that additional OREC 'peaks' were missed in the sampling. Visual inspection of the OREC-rich samples indicated that these samples had a distinctive greyish colour. Inspection of the core itself revealed no additional 'greyish sediments' which were not sampled.

The OREC peaks occurred solely during periods when climate was changing from interglacial to glacial. In sub-Saharan Africa, previous interglacials are thought to have been characterized by warm, wet conditions and the northward expansion of forest over savannah regions, whereas glacials are characterized by comparatively cool, dry conditions and the southward expansion of desert^{22,24}. Neither is conducive to high fire incidence. Large shifts in wind patterns, and hence source regions, could also produce large changes in OREC flux to the site, but, as discussed above, the absence of large changes in dust flux argues against such an interpretation.

These results differ dramatically from those of Verardo and Ruddiman¹⁷, who have presented a 200-kyr charcoal record for a core from the eastern equatorial Atlantic (RC24-07), 2,000 km southeast of ODP-668B. Their results suggests that charcoal abundances peaked in glacial times, with charcoal comprising up to 90% of the total (1–2%) organic carbon in the samples. These results are questionable on two grounds.

First, such a high proportion of elemental carbon in the samples should lead to a dramatic increase in the C/N ratio of organic carbon in the samples, possibly as high as 50 or more. The C/N ratios presented for the same core²⁷ generally range from 5 to 14,

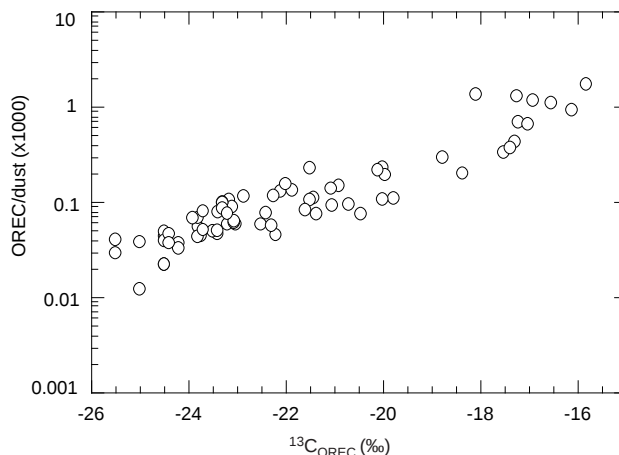


Figure 2 Relationship between the $\delta^{13}\text{C}$ value of oxidation-resistant elemental carbon (OREC) and its abundance as a ratio of the aeolian dust component in ODP-668B.

with occasional peaks up to 17. Furthermore, the peak in charcoal abundance inferred by Verardo and Ruddiman¹⁷ at the Last Glacial Maximum corresponds to samples with a very small range of C/N ratios between 8 and 12, suggesting that these samples are predominantly composed of marine-derived carbon. Isolated peaks in C/N ratio up to 18 in some samples may correspond to events similar to those identified in ODP-668B.

Second, the flux rates of elemental carbon calculated by Verardo and Ruddiman¹⁷ range as high as $50,000 \mu\text{g cm}^{-1} \text{kyr}^{-1}$, which is two orders of magnitude higher than calculated for a range of recent, Quaternary and Tertiary marine sediments by other researchers. Smith *et al.*¹⁶ calculated an average pre-industrial flux rate for Pacific and Atlantic sediments of $100 \mu\text{g cm}^{-1} \text{kyr}^{-1}$, whereas Herring¹¹ calculated Quaternary fluxes of $3\text{--}600 \mu\text{g cm}^{-1} \text{kyr}^{-1}$ and Tertiary fluxes of $0.3\text{--}10 \mu\text{g cm}^{-1} \text{kyr}^{-1}$ for Pacific Ocean sediments remote from coastal regions. These values agree well with the OREC fluxes calculated for ODP-668B ($1\text{--}400 \mu\text{g cm}^{-1} \text{kyr}^{-1}$), despite the fact that the data were generated by different techniques.

A sample of Antarctic marine sediment previously shown by Bird and Gröcke²⁸ to contain <1% elemental carbon was analysed in triplicate using the Verardo and Ruddiman¹⁷ technique. The results indicate that only $49 \pm 5\%$ of the carbon in this sample was removed by *in situ* nitric acid oxidation, implying an elemental carbon content of around 50%. It is possible that, whereas some organic carbon is oxidized directly to CO_2 , a considerable quantity of more refractory organic carbon is only partly oxidized and solubilized by the treatment but not removed, leading to an overestimate of 'charcoal' abundance using the Verardo and Ruddiman technique.

The OREC results from our study suggest that during the transition from interglacial to glacial mode the regional climate may have been destabilized and highly variable, leading to the build-up of fuel loads during wet periods followed by intense biomass burning in subsequent dry periods. In addition, biomass burning may be one mechanism by which high terrestrial organic carbon stocks built up during interglacials are shed at the onset of the subsequent glaciation.

The most recent peak in OREC abundance is unique in the past million years, in that it has occurred during an interglacial period. Anthropogenic biomass burning is considered the likely cause of this peak. Given that the current conditions of interglacial aridity in the region are also considered by some to be unusual^{5–7,22}, further consideration should be given to the possibility that a range of human activities in sub-Saharan Africa may have resulted in regional modifications to 'natural' vegetation patterns and climate. □

Received 22 July 1997; accepted 9 June 1998.

- Xue, Y. The influence and mechanisms of biosphere feedback on African climate. *Int. Ass. Hydrol. Sci. Publ.* **214**, 119–125 (1993).
- Hao, W. M., Liu, M.-H. & Crutzen, P. J. in *Fire in the Tropical Biota* (ed. Goldammer, J. G.) 452–465 (Ecological Studies Monogr. 184, Springer, Berlin, 1990).
- Andreae, M. O. in *Global Biomass Burning: Atmospheric, Climatic and Biospheric Implications* (ed. Levine, J. S.) 1–23 (MIT Press, Cambridge, MA, 1991).
- Bird, M. Fire, prehistoric humanity and the environment. *Interdisc. Sci. Rev.* **20**, 131–154 (1995).
- Kammen, D. M. & Marino, B. D. On the origin and magnitude of pre-industrial anthropogenic CO_2 and CH_4 emissions. *Chemosphere* **26**, 69–77 (1993).
- Schulz, E. Human impact and the anthropogenic nature of the west African Sahel. *Progr. Abstr. Inter-INQUA (Canberra, Australia)* **73** (Spec. Publ. Austral. Quat. Ass., Canberra, 1993).
- Hulme, M. & Kelly, M. Exploring the links between desertification and climate change. *Environment* **35**, 5–45 (1993).
- Petit-Maire, N. Lire l'avenir dans les archives Géologiques. *Recherche* **23**, 556–564 (1992).
- Rowell, D. P. *et al.* Modelling the influence of global sea surface temperatures on the variability and predictability of seasonal Sahel rainfall. *Geophys. Res. Lett.* **9**, 905–908 (1992).
- Smith, D. M., Griffin, J. J. & Goldberg, E. D. Spectrometric method for the quantitative determination of elemental carbon. *Anal. Chem.* **47**, 233–238 (1975).
- Herring, J. R. in *The Carbon Cycle and Atmospheric CO_2 : Natural Variations from Archean to Present* 419–442 (Geophys. Monogr. 32, Am. Geophys. Un., Washington DC, 1985).
- Wolbach, W. S. & Anders, E. Elemental carbon in sediments: Determination and isotopic analysis in the presence of kerogen. *Geochim. Cosmochim. Acta* **53**, 1637–1647 (1989).
- Emiliani, C., Price, D. A. & Seipp, J. in *Stable-Isotope Geochemistry: A tribute to Samuel Epstein* (eds Taylor, H. P., O'Neil, J. R. & Kaplan, I. R.) 229–231 (Spec. Publ. No. 3, Geochemical Soc., San Antonio, 1991).
- Verardo, D. J. Charcoal analysis in marine sediments. *Limnol. Oceanogr.* **42**, 192–196 (1997).
- Wolbach, W. S. *et al.* Global fire at the Cretaceous–Tertiary boundary. *Nature* **334**, 665–669 (1988).
- Smith, D. M., Griffin, J. J. & Goldberg, E. D. Elemental carbon in marine sediments: a baseline for burning. *Nature* **241**, 268–270 (1973).

- Verardo, D. J. & Ruddiman, W. F. Late Pleistocene charcoal in tropical Atlantic deep-sea sediments: climatic and geochemical significance. *Geology* **24**, 855–857 (1996).
- Cachier, H., Bremond, M. P. & Buat-Menard, P. Carbonaceous aerosols from different tropical biomass burning sources. *Nature* **340**, 371–373 (1989).
- Smith, B. N. & Epstein, S. Two categories of $^{13}\text{C}/^{12}\text{C}$ ratios for plants. *Plant Physiol.* **47**, 380–384 (1971).
- Ruddiman, W. F. *et al.* Site 668. *Proc. Init. Rep. ODP Leg 108 (pt 2)*, 931–951 (1988).
- Parkin, D. W., Phillips, D. R. & Sullivan, R. A. L. Airborne dust collection over the North Atlantic. *J. Geophys. Res.* **75**, 1782–1793 (1970).
- Pokras, E. M. & Mix, A. C. Eolian evidence for spatial variability of Late Quaternary climates in tropical Africa. *Quat. Res.* **24**, 137–149 (1985).
- deMenocal, P. B. Plio-Pleistocene African climate. *Science* **270**, 53–59 (1995).
- Dupont, L. Vegetation zones in NW Africa during the Brunhes Chron reconstructed from marine palynological data. *Quat. Sci. Rev.* **12**, 189–202 (1993).
- Sarnthein, M. *et al.* Stable-isotope stratigraphy for the last 750,000 years: "Meteor" core 13519 from the eastern equatorial Atlantic. *Meteor. Forsch. C* **38**, 9–24 (1984).
- Müller, P. J., Erlenkeuser, H. & von Grafenstein, R. in *Coastal Upwelling: its Sedimentary Record* (eds Theide, J. & Suess, E.) 365–398 (Plenum, New York, 1981).
- Verardo, D. J. & McIntyre, A. Production and destruction: control of biogenous sedimentation in the tropical Atlantic 0–300,000 years B.P. *Paleoceanography* **9**, 63–86 (1994).
- Bird, M. I. & Gröcke, D. Determination of the abundance and carbon-isotope composition of elemental carbon in sediments. *Geochim. Cosmochim. Acta* **61**, 3413–3423 (1997).

Acknowledgements. We thank P. Weiss and C. Mato for assistance with the provision of samples from ODP-668B. This work was supported by a Queen Elizabeth II fellowship to M.I.B. from the Australian Research Council.

Correspondence and requests for materials should be addressed to M.I.B. (e-mail: michael.bird@anu.edu.au).

Diachronous uplift of the Tibetan plateau starting 40 Myr ago

Sun-Lin Chung*, Ching-Hua Lo*, Tung-Yi Lee†, Yuquan Zhang‡, Yingwen Xie‡, Xianhua Li‡, Kuo-Lung Wang* & Pei-Ling Wang*

* Department of Geology, National Taiwan University, Taipei, Taiwan

† Department of Earth Sciences, National Taiwan Normal University, Taipei, Taiwan

‡ Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou, Guangdong

The uplift of the Tibetan plateau is generally regarded as a response to the convective removal of the lower portion of the thickened Asian lithosphere¹. This removal is also thought to be responsible for the east–west extension² that took place during the India–Asia collision. The timing of these events has been a subject of great interest for understanding mountain-building processes, collisional tectonics and the influence of these processes on climate change^{3,4}. In western Tibet, potassic lavas related to east–west extension were found to have been extruded over the past 20 Myr (refs 5, 6). Here we report the widespread occurrence of magmas in eastern Tibet which show similar geochemical signatures to the potassic lavas to the west but formed 40–30 Myr ago. These magmatic activities suggest a diachronous uplift history for the Tibetan plateau, with the convective removal of the lower lithosphere inducing rapid uplift in the east beginning some 40 Myr ago and in the west about 20 Myr later. This observation is consistent with sedimentation records from the Ganges–Brahmaputra delta to the Bengal fan^{7,8} and can better account for the tectonically driven models for strontium isotope evolution in the ocean⁹ and global cooling¹⁰ over the past 40 Myr.

The collision of India with Asia since early Cenozoic time has created the Tibetan plateau. The widespread occurrence of north–south-striking normal faults in Tibet (Fig. 1), starting before 14 Myr ago², suggests a switch of the tectonic regime from north–south convergence to east–west extension that could have resulted from convective removal of the lower part of lithospheric mantle by hotter and lighter asthenosphere¹. The latter process may furthermore have caused a sudden increase in the surface elevation and

perturbation of the geotherm, and thus induced potassic melt generation via small-degree melting of remnant lithospheric mantle if it had been previously enriched. Geochronological studies of movement of the normal faults², emplacement of the potassic magmas^{5,6}, and unroofing of basement rocks^{3,11}, have provided important constraints on the geodynamic evolution of the plateau. A widely accepted interpretation is that the rapid uplift of Tibet began around 20–13 Myr ago (refs 2–6). It is, however, based exclusively on information from the western part of the Tibetan plateau, that is, from regions located west of Lhasa or west of about 92° E (Fig. 1).

Compared with the investigations of western Tibet that form the present basis for understanding the Tibetan evolution, one of the few results from eastern Tibet is the discovery of late Palaeogene alkaline magmatic rocks¹². The rocks, divided here into five groups, are exposed around three tectonic belts, namely, the northern Ailao Shan-Red River shear zone, the Batang-Litang fault system and the Jinsha suture (Fig. 1). These magmas, however, were poorly known and their geodynamic significance remains unrecognized. A trachyte plug from the westernmost part of the magma suite (that is, from group 5) was the only outcrop documented in international publications¹³. The alkaline rocks, emplaced ~30 Myr ago^{12,14}, occur as small extrusive and intrusive bodies and consist of mafic to felsic lithologies. To the south, in eastern parts of the Lhasa block and the

Himalayas, contemporaneous igneous rocks (group 6) with similar petrological and geochemical characteristics have been reported in several localities¹⁵. Their easternmost counterpart crops out in the China–Vietnam border, where late Palaeogene potassic rocks have been transported from the eastern margin of Tibet by the mid-Tertiary sinistral movement along the Ailao Shan-Red River shear zone¹⁶. Therefore, the late Palaeogene alkaline rocks are considered widespread in the entire region of eastern Tibet.

Here we present Ar–Ar and K–Ar dating results of a variety of the alkaline rocks from groups 1–5 in eastern Tibet, and define a magmatic duration of 40–30 Myr ago (Table 1 and Fig. 2). Geochemically, these rocks range from basaltic to trachytic and rhyolitic composition (Fig. 3a). They show ultra-potassic or shoshonitic character that is comparable to Neogene (<20 Myr old) potassic volcanics^{5,6} emplaced in the western part of the Tibetan plateau (Fig. 3b). The similarity can also be shown by trace-element systematics of mafic rocks from the two magma suites of different ages. They both have incompatible trace-element variation patterns, with highly enriched large-ion lithophile elements (Rb, Ba, Th and U), light rare-earth elements and lead, and depleted high-field-strength elements, that is, Nb, Ta, Zr, Hf and Ti (Fig. 3c). Such potassic magmas, particularly those with high MgO contents (>10 wt%), require a metasomatized, phlogopite-bearing peridotite source residing in the lithospheric mantle (see ref. 6 for details).

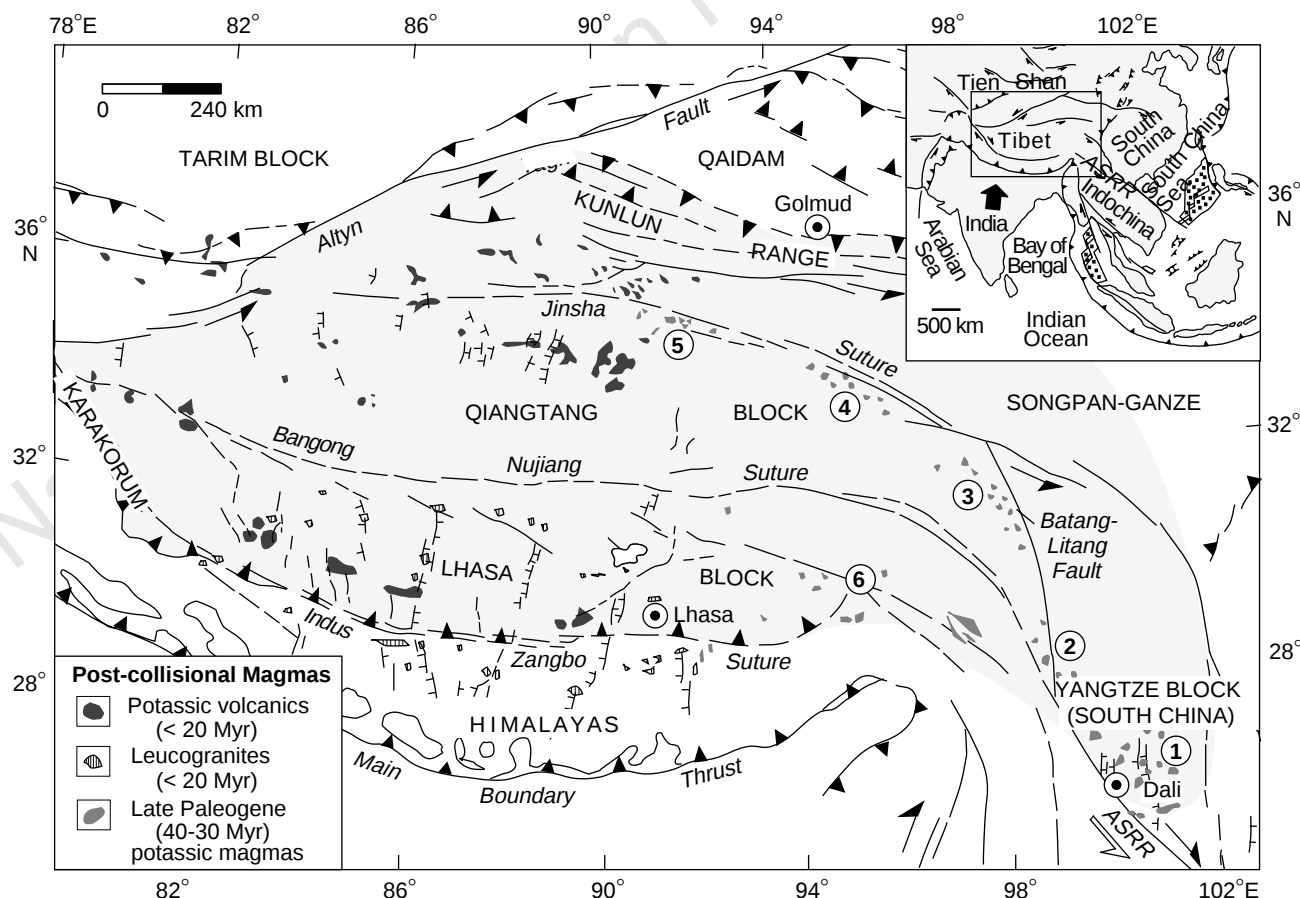


Figure 1 Simplified tectonic map of area around the Tibetan plateau (shaded area). Shown are the main tectonic blocks, structures, north-south-trending normal faults (after ref. 2), and post-collisional magmatic rocks (after refs 6, 12, 16 and 34). Inset shows tectonic units of southeast Asia related to the India–Asia collision (after ref. 28). The late Palaeogene, 40–30 Myr old, potassic magmas emplaced in eastern Tibet, east of 92° E, are classified into six groups. In this study, ~100 potassic rocks were collected from groups 1–5, and 40 samples were dated (see Table 1 and Fig. 2). We note that the restricted sampling is due to the

locations of river valleys and road cuts along the Jinsha River area, which were the only places accessible in this mountainous region. Although samples from group 6 were unavailable, their easternmost counterpart has been identified in the Vietnam–China border¹⁶ because of the 600-km sinistral block movement along the Ailao Shan-Red River (ASRR) shear zone. Samples from the latter have similar lithologies and chemical compositions and are synchronous with those from groups 1–5, so are considered co-genetic.

Changes in the tectonic setting are often recorded by changes in composition of the magmas formed. Eruption of the widespread, though small-volume, Neogene potassic lavas in western Tibet (Fig. 1) has been proposed to constrain the timing of the Tibetan uplift and lithospheric thinning in the region^{5,6}. The geothermal structure required to trigger melting in the lithospheric mantle could have been obtained only if the lower part of thickened Asian lithosphere was convectively removed and replaced by the asthenosphere^{1,5,6}. In eastern Tibet, the late Palaeogene potassic rocks from groups 1–5 are seemingly confined to three tectonic belts. However, the Jinsha suture represents an early Mesozoic suture and was inactive in the Cenozoic^{15,17}, and both the Ailao Shan–Red River and Batang–Litang fault systems display prominent strike-slip components active in the late Cenozoic^{16–18}. Motions of these tectonic belts could not have been responsible for generation

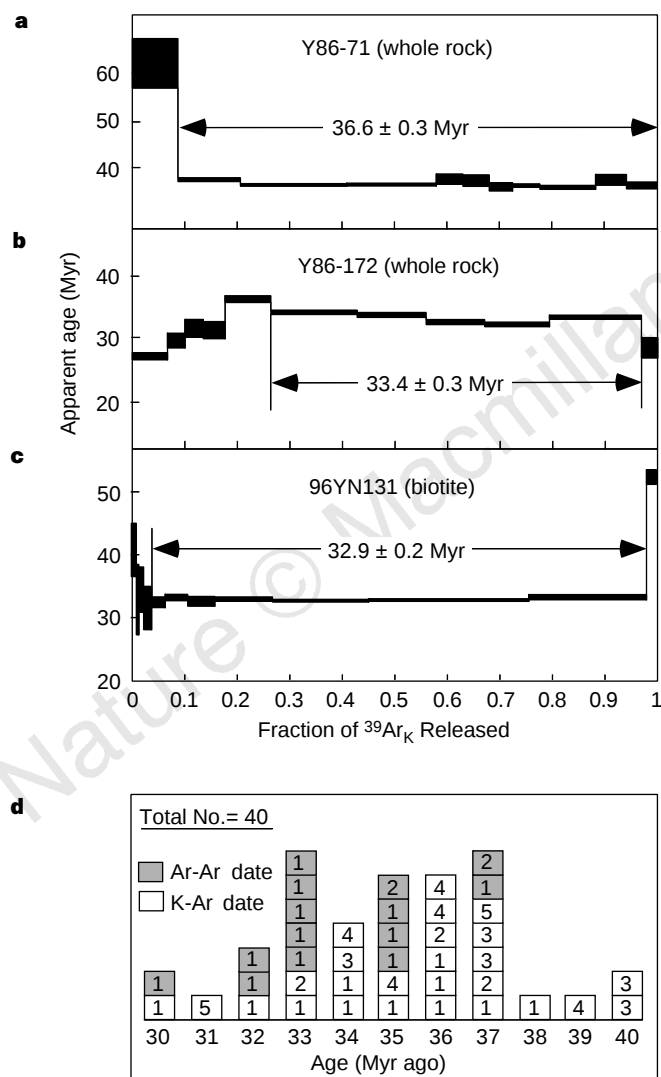


Figure 2 Representative ^{40}Ar – ^{39}Ar radiometric dating spectra. Shown are data for two whole rocks, Y86-71 (**a**) and Y86-172 (**b**), and a biotite separate, 96YN-13 (**c**), of potassic rocks from eastern Tibet. Plateau dates are shown with one standard error between arrows to indicate gas fractions included in age calculations. In **d**, a histogram of a total of 40 dates obtained by K–Ar and Ar–Ar methods is shown, thus defining a magmatic duration of 40–30 Myr ago. In addition to 26 K–Ar results covering five groups of sample locations (shown with numbers), the Ar–Ar age determinations were done for eight whole rocks and six mineral separates from groups 1 and 2. Detailed Ar–Ar age data and their experimental conditions, and major-element compositions of the samples dated, are available; see Supplementary Information.

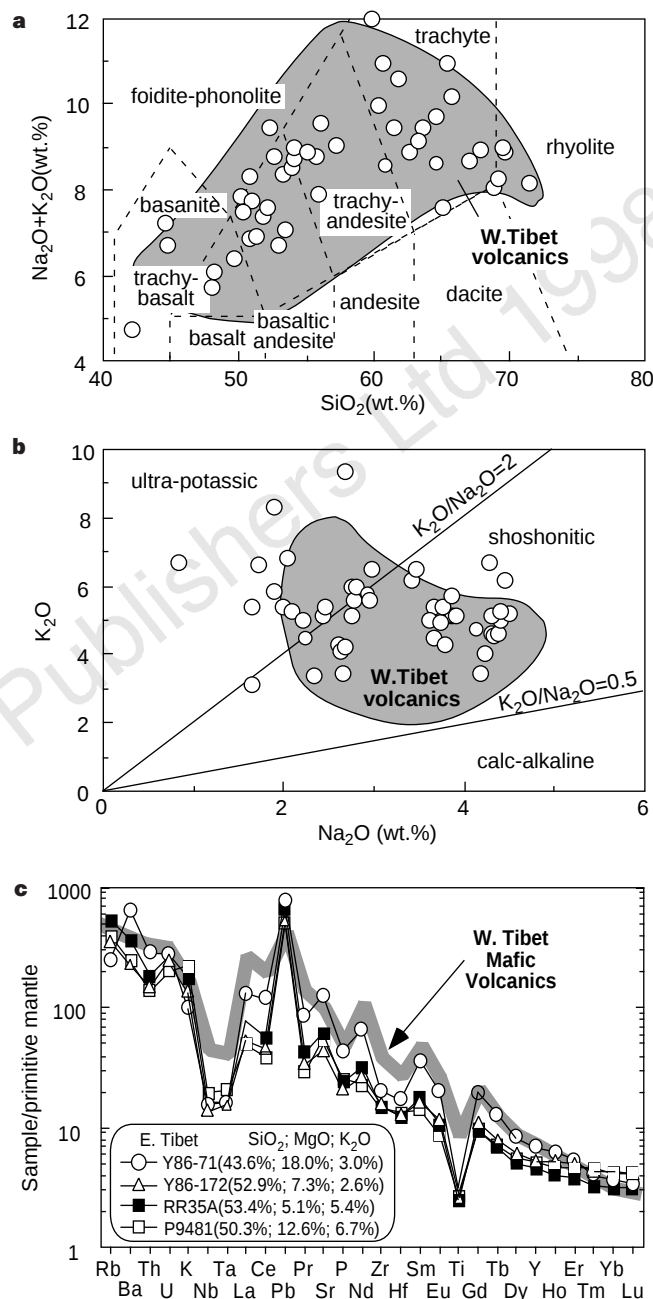


Figure 3 Comparison of compositions of rocks from eastern and western Tibet. Late Palaeogene magmatic rocks from eastern Tibet are plotted on **a**, total alkalis versus silica, **b**, potassium versus sodium, and **c**, primitive-mantle-normalized incompatible element variation diagrams, showing geochemical similarities with the Neogene potassic lavas from western Tibet. Data for the latter are from ref. 6. Classification schemes in **a** and **b** are according to ref. 35, and the normalizing values in **c** are from ref. 36. Major- and trace-element data plotted were obtained by X-ray fluorescence and dissolution inductively coupled plasma mass spectrometry methods at National Taiwan University and the Chinese Academy of Sciences, respectively, with the analytical uncertainties generally better than 5%. Likewise, both the potassic magma suites from Tibet have significantly lower neodymium and higher strontium isotopic ratios than those of the depleted convecting mantle. Combined elemental and isotopic data indicate that an enriched peridotite source containing phlogopite, which resides possibly within the lithospheric mantle, is required. The enrichment may have been related to certain ancient events or substantial amounts of sediment contamination by the Mesozoic Tethyan subduction zone processes (Y.Z., S.L.C., Y.X. and X.L., manuscript in preparation).

of the late Palaeogene, lithospheric-mantle-derived potassic melts. Thus, the occurrence of the two potassic magma suites suggests a diachronous lithospheric thinning beneath the Tibetan region. By analogy with formation of the younger suite associated with geothermal perturbation, normal faulting² and rapid increase in the elevation^{3,11} in western Tibet starting 20 Myr ago, the plateau may have undergone similar tectonomagmatic processes in its eastern part since as early as 40 Myr ago.

In comparison with the widely cited age of 50–45 Myr for the onset of the India–Asia collision, Beck *et al.*¹⁹ reported stratigraphic evidence for an earlier collision in northwest India before ~56 Myr ago. According to changes of the convergence rate derived from the northward motion of the Indian Ocean plate²⁰, a collision of India with Asia at the east end might have occurred since ~60 Myr ago. In this sense, the lithosphere underneath eastern Tibet could have reached its maximum sustainable thickness and thus being removed ~40 Myr ago. This could have caused a sudden uplift, regional extension and the potassic magmatism in eastern Tibet. Such

processes could have reached as far as western Yunnan in the South China block, as shown by the abundance of late Eocene to Oligocene rifted basins east of Diancangshan^{21,22}, northwest of Dali (Fig. 1), where potassic magmas of group 1 occur. Thus, western Yunnan at the eastern margin of Tibet consists of thinner crust (~40–45 km; ref. 23) and lower relief (~2 km on average²¹) relative to the rest of the plateau. It may be a consequence of “extensional collapse”, a process occurring after convective lithospheric thinning that in many cases has changed continental collision zones from mountains to basins²⁴.

Voluminous sediments related to the India–Asia collision have been accumulated in the Bengal fan (see ref. 8 for a review). Although the High Himalaya is generally regarded as an important source for sediments in the Bengal fan^{8,25}, a volume balance calculation²⁶ showed that the Himalayas cannot be the only sediment source and its hinterland (the Tibetan plateau) must have been significantly involved in shedding sediments before 20 Myr ago. Additionally, an onshore seismic stratigraphic analysis⁷, with

Table 1 Ar–Ar and K–Ar age determinations of potassic rocks from eastern Tibet

Ar–Ar age determinations*								
Sample	Location	Rock type	Phase dated	Plateau age (Myr ± σ)	Isochron age (Myr ± σ)	$(^{40}\text{Ar}/^{39}\text{Ar})_0$ (± σ)	MSWD	N
Y86-71	G-1	SH	WR	36.6 ± 0.3	30.5 ± 1.5	261.4 ± 13.0	3.4	11
Y86-172	G-1	TA	WR	33.4 ± 0.3	32.9 ± 0.3	288.2 ± 2.8	0.7	6
V111	G-1	TR	WR	35.3 ± 0.2	35.3 ± 0.2	304.8 ± 13.5	1.4	11
T770	G-1	TA	WR	35.1 ± 0.2	35.0 ± 0.2	335.1 ± 9.4	1.5	11
P201	G-1	SH	WR	35.3 ± 0.2	35.5 ± 0.4	269.5 ± 13.1	4.4	4
P9481	G-1	SH	WR	29.1 ± 0.9	30.0 ± 1.1	315.2 ± 18.6	6.8	13
YRR35A	G-2	TA	WR	36.9 ± 0.2	36.7 ± 0.3	319.0 ± 27.2	1.5	8
YRR35B	G-2	TA	WR	34.7 ± 0.2	34.7 ± 0.2	278.4 ± 12.3	1.3	9
96YN131	G-1	TA	Biot.	32.8 ± 0.2	32.8 ± 0.2	298.0 ± 16.9	0.5	7
96YN132	G-1	TA	Biot.	33.9 ± 0.3	33.9 ± 0.3	288.4 ± 2.9	2.1	11
96YN135	G-1	TA	Biot.	33.1 ± 0.2	33.1 ± 0.2	332.9 ± 18.8	0.9	8
96YN137	G-1	TA	Biot.	34.1 ± 0.3	34.1 ± 0.3	297.4 ± 6.1	1.7	11
96YN140	G-1	SH	Phlog.	32.9 ± 0.2	32.9 ± 0.2	294.7 ± 5.9	1.3	7
96YN148	G-1	SH	Phlog.	33.2 ± 0.2	33.2 ± 0.2	309.9 ± 26.1	0.7	8
K–Ar age determinations†								
Sample	Location	Rock type	Phase dated	K (wt%)	^{40}Ar (10^6 mol g^{-1})	Age (Myr ± σ)		
Y86-155	G-1	SH	WR	4.57	0.24349	30.4 ± 0.5		
83-703	G-1	GS	Biot.	7.93	0.48761	35.1 ± 0.5		
Y86-110	G-1	TR	WR	4.57	0.29922	37.3 ± 0.5		
Y86-130	G-1	SH	WR	1.93	0.11497	34.0 ± 0.5		
Y86-71	G-1	SH	Phlog.	7.19	0.45648	36.2 ± 0.5		
Y86-99	G-1	SH	Phlog.	7.51	0.46951	35.6 ± 0.5		
Y86-181	G-1	SH	Phlog.	7.39	0.46546	35.9 ± 0.5		
Y86-172	G-1	TR	WR	2.84	0.16122	32.4 ± 0.5		
Y86-185	G-1	SH	WR	5.85	0.33783	32.9 ± 0.5		
81-974	G-1	TR	Biot.	7.20	0.42285	33.5 ± 1.0		
81-751	G-1	SY	Biot.	7.40	0.49374	38.0 ± 0.5		
81-862	G-2	TR	Biot.	7.30	0.46999	36.7 ± 0.6		
81-599	G-2	SY	Biot.	7.28	0.46254	36.2 ± 0.8		
81-911	G-2	SY	Biot.	7.25	0.43509	33.0 ± 0.7		
83-55	G-3	TR	Biot.	7.63	0.49852	37.2 ± 0.6		
83-105	G-3	SY	K-feld.	7.34	0.43984	34.2 ± 0.6		
83-99	G-3	TR	Biot.	7.56	0.49200	37.1 ± 0.6		
83-931	G-3	GS	Biot.	7.64	0.52970	39.5 ± 0.7		
82-106	G-3	GS	Biot.	7.39	0.53109	41.0 ± 0.7		
N-1	G-4	TR	Biot.	6.79	0.41291	34.1 ± 1.0		
N-2	G-4	RH	WR	3.72	0.23023	35.8 ± 1.0		
N-3	G-4	GS	Biot.	7.49	0.52528	39.2 ± 1.0		
ZH-1	G-4	GS	WR	4.57	0.27753	35.0 ± 1.0		
ZA-1	G-4	GS	K-feld.	8.82	0.54680	35.5 ± 1.0		
85-142	G-5	TR	Biot.	7.21	0.39620	31.4 ± 0.5		
EDG-1	G-5	TR	WR	2.80	0.07300	37.3 ± 1.0		

Location: symbols G-1 to G-5 indicate samples from groups 1 to 5 as shown in Fig. 1. Rock type: SH, shoshonite; TA, trachy-andesite; TR, trachyte; GS, granosyenite; SY, syenite; RH, rhyolite. Phase dated: WR, whole rock; Biot., biotite; phlog., phlogopite; K-feld., potash feldspar.

* $(^{40}\text{Ar}/^{39}\text{Ar})_0$ is the ratio at intercept, MSWD is the mean squared weighted deviate and N is the number of steps used in age calculation. A plateau was defined by at least 60% of released ^{39}Ar at least four consecutive steps, and if the integrated plateau age agreed with each individual age of steps forming the plateau within a confidence level of 2σ . Samples were wrapped in aluminium foil packets and irradiated for 8 h in the THOR nuclear reactor at Sing-Hua University, Taiwan. Maximum flux gradient was 0.9% within overall sample volume. The irradiated samples were step-heated using a Lindberg resistance furnace and analysed by a Varian-MAT GD150 mass spectrometer at National Taiwan University. The monitor standard was the LP-6 biotite (127.7 ± 1.4 Myr) and ages were calculated using the constants recommended by Steiger and Jäger²¹. All errors are quoted at the 1σ level and do not include the uncertainty of the monitor age. Experimental details as given in refs 32, 33.

† K–Ar age data were determined by conventional potassium–argon dating procedures at Institute of Geochemistry, Chinese Academy of Sciences. Uncertainties given as 1σ level.

accompanying well data, in the Bengal basin has revealed that the Ganges–Brahmaputra delta started to grow rapidly ~40 Myr ago with an increase of sediment flux and development of significant prograding clastic depositional sequences. These sedimentation records are consistent with our view (based on the magmatism) that suggests an earlier rapid rise of eastern Tibet, from where thick piles of the lower sedimentary sequences in the Ganges–Brahmaputra delta and the Bengal fan could have been derived. A similar diachronous uplift history has been also proposed in the northern margin of the Tibetan plateau²⁷. On the basis of radiometric dating results of granites and gneisses, Arnaud *et al.*²⁷ have argued that the basement unroofing and uplift in the eastern Kunlun Range occurred since ~35 Myr ago, and apparently preceded those in the western part of the Range.

Consequently, we suggest that the Tibetan plateau has undergone two main stages of rapid uplift caused by diachronous removal of thickened Asian lithosphere after the Indian indentation. Whereas the younger, and better-known, uplift began ~20 Myr ago in the western part of Tibet, the earlier event took place in the east, since ~40 Myr ago. Our observation can be reconciled with tectonic forcing models for the Cenozoic isotope evolution in the ocean⁹ and global climate change¹⁰; these models put forward the effects of the rise and subsequent erosion and weathering of the Tibetan–Himalayan region. More specifically, it provides the first (to our knowledge) convincing time constraint that accommodates not only the rapid and steady increase in the seawater strontium isotope ratios beginning 40 Myr ago⁹ but also the onset of global cooling from the early Eocene¹⁰.

We further propose that the late Palaeogene tectonic processes in eastern Tibet played an important role in initiating the Ailao Shan–Red River shear zone, thus causing the continental extrusion²⁸ which eventually offset the Indochina block from the South China block for about 600 km (ref. 16). This may be due to the occurrence of a mechanically weakened lithosphere domain in the eastern Tibetan region as a result of the earlier lithospheric removal associated with extension and upwelling of hotter asthenosphere. The continental extrusion took place from about 27 to 17 Myr ago^{16,29}, nearly coincident with the motion of the Gangdese thrusting system in the southernmost Tibet^{3,30}. Such a synchronicity relates to the long-standing debate on the relative importance of crustal thickening and lateral extrusion after the India–Asia collision. Future detailed investigations of this inaccessible mountainous region are needed for a more comprehensive understanding of the geological evolution of Tibet. In the absence of such data, any tectonic approaches using the one-stage uplift model for the entire Tibetan plateau may have to face the problem of oversimplification. □

Received 22 April 1997; accepted 10 June 1998.

- Platt, J. P. & England, P. C. Convective removal of lithosphere beneath mountain belts: Thermal and mechanical consequences. *Am. J. Sci.* **294**, 307–336 (1994).
- Coleman, M. & Hodges, K. Evidence for Tibetan plateau uplift before 14 Myr ago from a new minimum age for east-west extension. *Nature* **374**, 49–52 (1995).
- Harrison, T. M., Copeland, P., Kidd, W. S. F. & Yin, A. Raising Tibet. *Science* **255**, 1663–1670 (1992).
- Molnar, P., England, P. & Martinod, J. Mantle dynamics, the uplift of the Tibetan plateau, and the Indian monsoon. *Rev. Geophys.* **31**, 357–396 (1993).
- Turner, S. *et al.* Timing of Tibetan uplift constrained by analysis of volcanic rocks. *Nature* **364**, 50–54 (1993).
- Turner, S. *et al.* Post-collisional, shoshonitic volcanism on the Tibetan plateau: implications for convective thinning of the lithosphere and the source of ocean island basalts. *J. Petrol.* **37**, 45–71 (1996).
- Lindsay, J. F., Holliday, D. W. & Hulbert, A. G. Sequence stratigraphy and the evolution of the Ganges–Brahmaputra Delta complex. *Am. Assoc. Petrol. Geol. Bull.* **75**, 1233–1254 (1991).
- Curry, J. R. Sediment volume and mass beneath the Bay of Bengal. *Earth Planet. Sci. Lett.* **125**, 371–383 (1994).
- Richter, F. R., Rowley, D. B. & DePaolo, D. J. Sr isotope evolution of seawater: The role of tectonics. *Earth Planet. Sci. Lett.* **109**, 11–23 (1992).
- Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359**, 117–122 (1992).
- Copeland, P., Harrison, T. M., Pan, Y., Kidd, W. S. F. & Roden, M. Thermal evolution of the Gangdese batholith, southern Tibet: A history of episodic unroofing. *Tectonics* **14**, 223–236 (1995).
- Zhang, Y. Q., Xie, Y. W. & Tu, G. Z. Preliminary studies of the alkali-rich intrusive rocks in the Ailaoshan–Jinshajiang belt and their bearing on rift tectonics. *Acta Petrol. Sinica* **3**, 17–26 (1987).
- Pearce, J. A. & Mei, H. Volcanic rocks of the 1985 Geotraverse: Lhasa to Golmud. *Phil. Trans. R. Soc. Lond. A* **327**, 169–201 (1988).

- Zhu, B. Q., Zhang, Y. Q. & Xie, Y. W. Isotope characteristics of Cenozoic potassic volcanic rocks from Haidong, Yunnan, and their implications for subcontinental mantle evolution in southwestern China. *Geochimica* **21**, 201–212 (1992).
- Bureau of Geology and Mineral Resources of Xizang Region *Regional Geology of Xizang (Tibet)* (Geol. Publ. House, Beijing, 1993).
- Chung, S. L. *et al.* Intraplate extension prior to continental extrusion along the Ailao Shan–Red River shear zone. *Geology* **25**, 311–314 (1997).
- Ratschbacher, L., Frisch, W., Chen, C. & Pan, G. In *The Tectonic Evolution of Asia* (eds Yin, A. & Harrison, T. M.) 227–249 (Cambridge Univ. Press, 1995).
- Holt, W. E., Ni, J. F., Wallace, T. C. & Haines, A. J. The active tectonics of the eastern Himalayan syntaxis and surrounding regions. *J. Geophys. Res.* **96**, 14595–14632 (1991).
- Beck, R. A. *et al.* Stratigraphic evidence for an early collision between northwest India and Asia. *Nature* **373**, 55–57 (1995).
- Lee, T. Y. & Lawver, L. A. Cenozoic plate reconstruction of Southeast Asia. *Tectonophysics* **251**, 85–138 (1995).
- Bureau of Geology and Mineral Resources of Yunnan Province *Regional Geology of Yunnan Province* (Geol. Publ. House, Beijing, 1990).
- Leloup, P. H. *et al.* The Ailao Shan–Red River shear zone (Yunnan, China), Tertiary transform boundary of Indochina. *Tectonophysics* **251**, 3–84 (1995).
- Kan, R. J., Hu, H. X., Zeng, R. S., Mooney, W. D. & McEvilly, T. V. Crustal structure of Yunnan Province, People's Republic of China, from seismic refraction profiles. *Science* **234**, 433–437 (1986).
- Houmear, G. From mountains to basin. *Nature* **379**, 771–772 (1996).
- Copeland, P. & Harrison, T. M. Episodic rapid uplift in the Himalaya revealed by ⁴⁰Ar/³⁹Ar analysis of detrital K-feldspar and muscovite. *Geology* **18**, 354–357 (1990).
- Johnson, M. R. W. Volume balance of erosional loss and sediment deposition related to Himalayan uplifts. *J. Geol. Soc. Lond.* **151**, 217–220 (1994).
- Arnaud, N. O. *et al.* Ages of magmatism and tectonism of north eastern Kunlun, China. *30th Int. Geol. Congr. (Beijing) Abstr.* **1**, 192 (1996).
- Trapponnier, P., Peltzer, G., Armijo, R., Le Dain, A. Y. & Cobbold, P. Propagating extrusion tectonics in Asia: New insights from simple experiments with plasticine. *Geology* **10**, 611–616 (1982).
- Wang, P. L. *et al.* Movement of the Ailao Shan–Red River shear zone: A Vietnamese perspective from thermochronological evidence. *Eos* **78**, F649 (1997).
- Yin, A. *et al.* Tertiary structural evolution of the Gangdese thrust system in southeastern Tibet. *J. Geophys. Res.* **99**, 18175–18201 (1994).
- Steiger, R. H. & Jäger, E. Submission on geochronology: Convention on the use of decay constants in geo- and cosmochronology. *Earth Planet. Sci. Lett.* **36**, 359–362 (1977).
- Lo, C. H. & Lee, C. Y. ⁴⁰Ar/³⁹Ar method of K–Ar age determination of geological samples using sing-Hua Open-pool (HOR) Reactor. *J. Geol. Soc. China* **37**, 143–164 (1994).
- Lo, C. H., Onstot, T., Chen, C. H. & Lee, T. Ar/Ar dating of andesitic volcanism in the Luzon arc near Taiwan. *Chem. Geol.* **114**, 157–178 (1994).
- Le Fort, P. Granite in the tectonic evolution of the Himalaya, Karakorum, and southern Tibet. *Phil. Trans. R. Soc. Lond. A* **327**, 281–299 (1988).
- Le Maitre, R. W. (ed.) *A Classification of Igneous Rocks and Glossary of Terms* (Blackwell Sci. Publ., Oxford, 1989).
- Sun, S.-s. & McDonough, W. F. In *Magmatism in the Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 313–345 (Spec. Publ. 42, Geol. Soc., London, 1989).

Supplementary information is available on Nature's World Wide Web side (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank A. Crawford, S.-s. Sun and F. Yang for suggestions on the earlier drafts, and N. Rogers and P. England for reviews which significantly improved this Letter. This study was supported by the National Science Council, Taiwan and the National Science Foundation, China.

Correspondence and requests for materials should be addressed to S.L.C. (e-mail: sunlin@ccms.ntu.edu.tw).

Evolutionary transition from stretch to hearing organs in ancient grasshoppers

Moir J. van Staaden* & Heiner Römer

Institute for Zoology, University of Graz, A-8010 Graz, Austria

Ears of modern insects occur on a wide variety of body parts and are thought to have evolved from ubiquitous stretch or vibration receptors^{1–4}. This relationship, based on comparative anatomy and similarities in the embryological development of ears in divergent taxa^{5–7}, has led to the widespread assumption of homology of these structures in insects, although this has not been tested rigorously. Here we report on the hearing organs of a relatively ancient⁸, atympanate bladder grasshopper^{9–11} (*Bullacris membracioides*), which is capable of signalling acoustically over ~2 km¹². We show that, within single individuals of this species, serially repeated abdominal ears show functional continuity from simple to more complex forms. All 12 morphologically differentiated

* Present address: Department of Biological Sciences, Bowling Green State University, Bowling Green, Ohio 43403, USA.

organs respond to sound frequencies and intensities that are biologically significant, and mediate adaptive behavioural responses. By linking observations at the anatomical, physiological and behavioural level, our experiments provide evidence for the transition in function and selective advantage during the evolutionary development of this complex structure^{13,14}. It is possible that ancestral insects with only simple pleural receptors had auditory capability covering distances substantially greater than contemporary insects with tympanate ears.

Although they lack differentiated tympana, bladder grasshoppers (Pneumoridae, Orthoptera) possess hearing organs homologous to those found in locusts. Histological staining shows that this organ contains nearly 2,000 sensory units (Fig. 1a). Multicellular scolopidia, each consisting of a bipolar sensory neuron and several accessory cells, are attached to the pleural cuticle of the first abdominal segment (A1) by means of two bundles of very long attachment cells (~1.4 mm in length compared with <100 µm in the tympanate hearing organs of modern grasshoppers). In addition, five pairs of smaller chordotonal organs form at the lateral body wall of abdominal segments A2–A6. Having only 11 scolopidia each, they attach by apical and basal cellular anchorages (Fig. 1b), with dendrites and attachment cells similar in length to those in A1.

We determined the neurophysiological function of both forms of abdominal chordotonal organ by recording extracellularly from nerve N1 of the respective ganglia. The male mating call is a resonant ($Q_{10\text{dB}} = 4.1\text{--}4.6$) call with a carrier frequency of ~1.7 kHz, comprising a resonant sixth syllable (98 dB SPL sound pressure level (SPL) at 1 m) and five lower-intensity (70 dB SPL) introductory syllables¹². We found that all abdominal chordotonal organs responded to acoustic stimulation within an intensity range that was biologically meaningful. Receptors in A1 have best frequencies of 4 kHz (Fig. 2), apparently mismatched to the 1.7 kHz carrier frequency of the male call^{15–17}. Despite this anomaly and the absence of an overt tympanum, the sensitivity of the A1 hearing organ at its best frequency is one of the highest recorded so far for insects (12.8 ± 4.89 dB SPL; $n = 5$) and, together with the intense calling songs, accounts for the extraordinarily large communication distance. In contrast, the tuning of receptors in segments A2–A6 matches the species-specific male signal (1.5–2 kHz) but they are

much less sensitive. Both individual tuning curves (Fig. 2) and mean threshold intensities (Fig. 3a) demonstrate increasing thresholds (58–77 dB SPL) from anterior to posterior segments. Nonetheless, the temporal song pattern of the male signal is detected in the spike discharge of all pleural chordotonal receptors and is faithfully represented when at suprathreshold intensity for a given segment (Fig. 3c; see A2 receptors).

It is the behaviour mediated by the serial chordotonal receptors, however, that establishes their role as functional hearing organs. First, we know from neurophysiological experiments with a biological microphone¹⁸ (that is, a small, portable set-up for recording action-potential activity of auditory neurons directly in the field) that females have an average neural sensitivity of 32 dB SPL to the male call, and hear males calling at a distance of 1–2 km. However, in behavioural playback experiments, females show no auditory response until stimulation intensity exceeds 60 dB SPL (Fig. 3a). This in itself is not unusual as behavioural thresholds frequently exceed neurophysiological thresholds by such margins¹⁹. More unusual is a conspicuous sexual asymmetry in both the active space and the degree of stereotypy of mate location calls. Whereas females hear the male call with their most sensitive hearing organ in A1 over distances up to 1.9 km, males hear the soft female response over a maximum of only 50 m. Sound pressure levels of 60 dB, as perceived by females, imply that males are 50–100 m distant. This distance approximates the active space of the female call for the male, indicating that females risk auditory exposure to nocturnal predators only when a potential mate is clearly within her acoustic transmission range. There would have been strong selection for chordotonal receptor mechanisms that discriminate between 'advantageous' and 'disadvantageous' airborne sound signals and such mechanisms should evolve to maximize the expected utility of the receiver's response²⁰ in this antiphonal duetting system.

Second, playback experiments show that this adaptive behavioural response is graded. In contrast to the stereotyped male call, the female response varies from one to eight syllables. Behavioural tests showed this to be a function of male call intensity; beyond the response threshold females add approximately one additional syllable for each 3-dB increase in male call intensity (Fig. 3b). We examined the neurophysiological basis for this behaviour by testing the differential responses of pleural chordo-

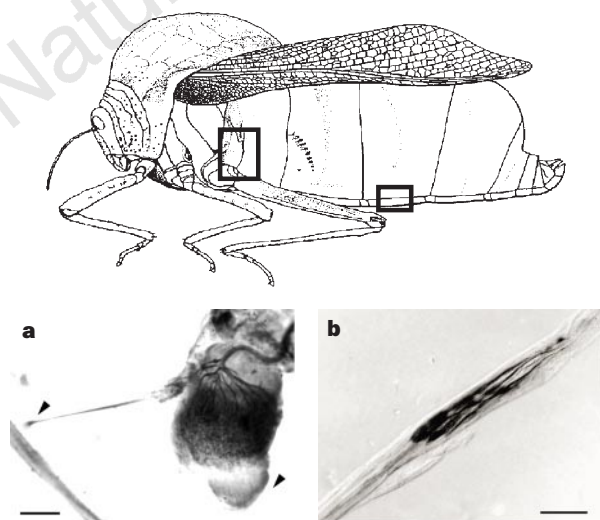


Figure 1 Chordotonal organs in abdominal segments A1–A6 of adult bladder grasshoppers are differentiated morphologically, exhibiting two levels of complexity. **a**, Cobalt backfill of A1 reveals two cuticular connections (arrowed) with 32 scolopidia in the smaller branch and ~2,000 in the larger. Scale bar, 300 µm. **b**, Phase-contrast microscopy of a chordotonal organ with stained sensory cells attached to the pleura in A3. Scale bar, 50 µm.

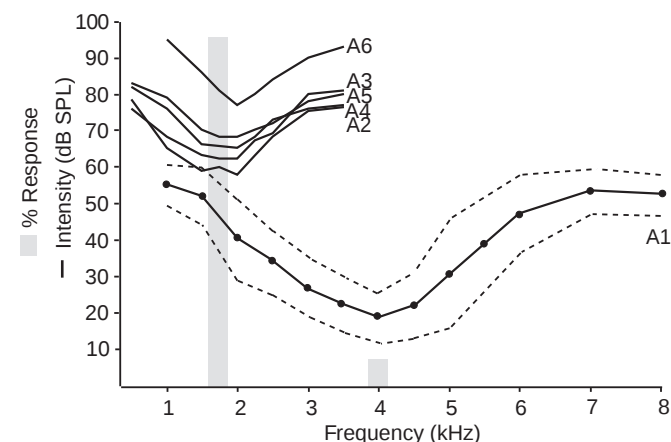


Figure 2 Six pairs of abdominal chordotonal organs in adult pneumorids respond to sound with action-potential discharges at frequencies and intensities that are biologically significant. Neurophysiological tuning curve of A1 is for males and females combined (dotted lines indicate standard deviation limits of the mean tuning curve; $n = 11$). The curves for A2–A6 are from a single individual and thus do not reflect the means presented in Fig. 3a. Shaded bars demonstrate reliable behavioural responses of adult females to pure tones of 1.7 kHz and the paucity of response at 4 kHz. Note that an SPL of 75 dB at 4 kHz may be suprathreshold for some pleural organs.

tonal organs in females using playbacks of the male call, which simulated sender–receiver distances from 5 m to ~100 m. As a consequence of the different thresholds of the organs in A2–A6, the action-potential discharge representing the temporal pattern of the male call varies between the serially repeated segments at a given distance (Fig. 3c). How far away the male is from the female is thus, at least partly, encoded in the output of serially repeated sensory organs and the female's response would be subject to strong directional selection.

We have shown that the serial receptors in A1–A6 of *B. membracioides* fulfil all three criteria for functional ears: morphological, neural and behavioural¹⁴. That they are all true ears is further supported by results from ablation experiments in which we removed the A1 organs bilaterally from previously responsive females. Subsequent neurophysiological experiments confirmed the total absence of neural activity in the A1 afferent nerves of these individuals, but bilaterally operated females continued to respond to male calls at distances from 8–16 m in playback experiments. In addition, behavioural playback experiments to intact animals provide more direct support for the role of A2–A6 in mediating the adaptive female response. By using presentations of pure tones of 1.7 kHz and 4 kHz, we found that females responded almost invariably and exclusively to the lower frequency (96% versus 8%, $n = 76$ presentations; Fig. 2). This was despite a stimulus intensity that was ~40 dB above the hearing threshold at 4 kHz for the organ in A1, and <15 dB above that at 1.7 kHz for the organs in A2–A5. The few responses observed at 4 kHz most likely result from suprathreshold activation of pleural organ receptors at this frequency, rather than of A1 receptors. Ancestral insects with only simple pleural receptors may likewise have had auditory capability over distances higher than those for effective substrate-borne vibration, and substantially greater than many contemporary insects with tympanate ears.

We believe that these results illustrate the transition in function and selective advantage occurring during the evolution of complex suites of interdependent characters. Several lines of evidence indicate that chordotonal hearing organs represent a transitional form rather than a specialization unique to *B. membracioides*, although

we do not suggest that the organs in *Bullacris* are themselves intermediate. Neurophysiological experiments on this species show that both chordotonal and auditory organs are sensitive to respiratory movements, highlighting their common origin from receptor cells involved in the perception of abdominal ventilatory movements^{21,22}. Despite continuing questions concerning the precise phylogenetic placement of the Pneumoridae within the Pneumoroidea, all members of the superfamily lack tympana and possess a stridulatory apparatus, but only bladder grasshoppers have evolved an inflated abdomen and conspicuous calling. Abdominal inflation *per se* is not essential for the operation of chordotonal hearing organs, as demonstrated by females and alternate males¹¹ that lack this feature but share the same two forms of functional ears. Comparative data from other Pneumoroidea are necessary to assess the significance of conspicuous signalling for the evolution of chordotonal organs with auditory function, but clearly the evolution of specialized tympana and auditory organs need not be closely correlated. Molecular phylogenetic analyses suggest that the Pneumoridae probably arose in the Jurassic, before the tympanate Pyrgomorphidae, and from ancestors in which external tympana also appear to be absent⁸. We have examined the more basal atympanate Proscopiidae and although we should be cautious in declaring the absence of a structure, to date we have found no evidence of either ear form. It is likely that the pneumorid lineage has never contained species with classical ears, and that the extant species have retained an ancient form of hearing that represents a transitional stage in the evolution of insect ears.

Developmental and comparative studies of eared and earless insects have demonstrated modularity in structure and serial homology of auditory and mechanosensory receptors, suggesting that tympanic receptors might be a special part of the general chordotonal system common to all insect groups. To this modularity in structure we now add that of function. We propose that the functional continuity between classical auditory and pleural chordotonal organs in *B. membracioides* typifies the evolutionary trajectory of transitional forms from ancestral proprioceptors to long-distance auditory receptors, which has occurred independently in divergent insect taxa. □

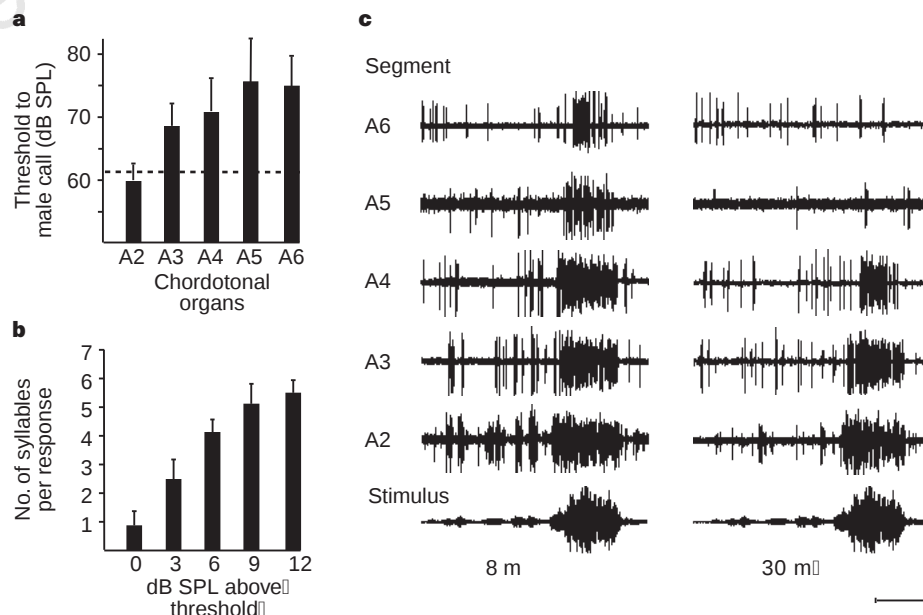


Figure 3 Chordotonal organs mediate adaptive behavioural responses to acoustic signals. **a**, Mean neurophysiological thresholds of organs A2–A6 of females ($n = 10$) to the male call. Note the close correspondence between the A2 threshold and the behavioural response threshold (dashed line). **b**, Behavioural response of females to playback of male calls that are between 0 and 12 dB above

response threshold. Females add approximately one syllable to their auditory response for each 3-dB increase in SPL. **c**, Differential neurophysiological responses of pleural chordotonal organs in females to playbacks of a male call simulating sender–receiver distances of 8 and 30 m in the field. Scale bar, 500 ms.

Methods

Neurophysiology. Experimental animals were collected at Inchanga (Kwa-Zulu/Natal, South Africa) and all field experiments conducted in February 1994–96 and March 1998. Numbers of sensory cells were determined either by backfilling the axons with cobalt lysine and intensifying the staining with silver according to standard methods²³ (for A2–A6), or by counting the number of attachment cells in 5- μ m cross-sections of the chordotonal organ stained with methylene blue (for A1). Extracellular recordings of the afferent activity of pleural chordotonal organ receptors were obtained with silver-wire hook electrodes attached to nerve N1 of the respective neuromeres of the metathoracic ganglion or abdominal ganglia. Tuning curves for A2–A6 receptors were obtained by stimulating the preparation with digitized, pure-tone sounds pulses (sampling rate 44 kHz; 50 ms duration) of variable frequency (0.5–8 kHz) and intensity. Tuning curves of A1 receptors were established in a sound-damped chamber in Graz as previously described²⁴. To determine how well the time pattern of the male signal is represented in the spike discharge of female chordotonal organs, we stimulated receptors in A2–A6 with digitized models of the male call (sampling rate 44 kHz) at SPLs ranging from 55 to 84 dB, thus simulating different sender–receiver distances.

Behaviour. Playback experiments were conducted as described¹². Chordotonal organ ablation was done on anaesthetized females by opening the body wall at the attachment site of the hearing organ in A1, removing the organ and sealing the opening with histoacryl. Control animals were similarly sham-treated. Bilaterally operated females were tested the following day using playbacks of the male call at various intensities. Threshold was defined as the SPL of the male call that elicited at least one female chirp in response to the male call in four out of five presentations. In all ablation experiments we subsequently confirmed the absence of neural activity in the afferent nerves of the receptors in A1 with neurophysiological methods. Behavioural tuning experiments were performed in a sound-damped room using digitized, pure-tone sound pulses (sampling rate 22 kHz; 880 ms duration) of 1.7 and 4 kHz, randomly presented at 75 dB SPL at the position of the female and intervals of 40 s.

Received 8 December 1997; accepted 16 June 1998.

- Boyan, G. S. Another look at insect audition: the tympanic receptors as an evolutionary specialization of the chordotonal system. *J. Insect Physiol.* **39**, 187–200 (1993).
- Fullard, J. H. & Yack, J. E. The evolutionary biology of insect hearing. *Trends Ecol. Evol.* **8**, 248–252 (1993).
- Hoy, R. R. & Robert, D. Tympanal hearing in insects. *Annu. Rev. Entomol.* **41**, 433–450 (1996).
- Shaw, S. R. Detection of airborne sound by a cockroach 'vibration detector': a possible missing link in insect auditory evolution. *J. Exp. Biol.* **193**, 13–47 (1994).
- Meier, T. & Reichert, H. Embryonic development and evolutionary origin of the Orthopteran auditory organs. *J. Neurobiol.* **21**, 592–610 (1990).
- Yack, J. E. & Rösöts, B. I. The metathoracic wing-hinge chordotonal organ of an atympanate moth *Actias luna* (Lepidoptera, Saturniidae): a light- and electron-microscopic study. *Cell Tissue Res.* **267**, 455–471 (1992).
- Yack, J. E. & Fullard, J. F. The mechanoreceptive origin of insect tympanal organs: a comparative study of similar nerves in tympanate and atympanate moths. *J. Comp. Neurol.* **300**, 523–534 (1990).
- Flook, P. K. & Rowell, C. H. F. The phylogeny of the Caelifera (Insecta, Orthoptera) as deduced from mtDNA gene sequences. *Mol. Phylogenet. Evol.* **8**, 89–103 (1997).
- Dirsh, V. M. Revision of the Family Pneumoridae (Orthoptera: Acridoidea). *Bull. Br. Mus. Nat. Hist. Entomol.* **15**, 325–396 (1965).
- Alexander, A. J. The bladder grasshopper: a 'nu-nu' of mystery and intrigue. *Afr. Wildl.* **46**, 261–262 (1992).
- Alexander, A. & van Staaden, M. J. In *Alternative Life-History Styles of Animals* (ed. Bruton, M. N.) 261–277 (Kluwer, Dordrecht, 1989).
- van Staaden, M. J. & Römer, H. Sexual signalling in bladder grasshoppers: tactical design for maximizing calling range. *J. Exp. Biol.* **200**, 2597–2608 (1997).
- Robert, D., Amoroso, J. & Hoy, R. R. The evolutionary convergence of hearing in a parasitoid fly and its cricket host. *Science* **258**, 1135–1137 (1992).
- Yack, J. E. & Fullard, J. F. What is an insect ear? *Ann. Entomol. Soc. Am.* **86**, 677–682 (1993).
- Popov, A. V. Sound production and hearing in the cicada *Cicadetta sinuatifemoris* Osh. (Homoptera, Cicadidae). *J. Comp. Physiol. A* **142**, 271–280 (1981).
- Bailey, W. J. & Römer, H. Sexual differences in auditory sensitivity: mismatch of hearing threshold and call frequency in a tettigoniid (Orthoptera, Tettigoniidae: Zaprochilinae). *J. Comp. Physiol. A* **169**, 349–353 (1991).
- Mason, A. C. Hearing in a primitive ensiferan: the auditory system of *Cyphoderris monstrosa* (Orthoptera: Haglidae). *J. Comp. Physiol. A* **168**, 351–363 (1991).
- Rheinlaender, J. & Römer, H. Insect hearing in the field I. The use of identified nerve cells as 'biological microphones'. *J. Comp. Physiol. A* **158**, 647–651 (1986).
- Yager, D. D. & Spangler, H. G. Behavioural response to ultrasound by the tiger beetle *Cincindela marutha* does combine aerodynamic changes and sound production. *J. Exp. Biol.* **200**, 649–659 (1997).
- Wiley, R. H. In *Behavioral Mechanisms in Ecology* (ed. Real, L. A.) 157–189 (Univ. Chicago Press, 1994).
- Hustert, R. Morphologie und Atmungsbewegungen des 5. Abdominalsegments von *Locusta migratoria migratorioides*. *Zool. Jb. Physiol.* **78**, 157–174 (1974).
- Meyer, J. & Hedwig, B. The influence of tracheal pressure changes on the responses of the tympanal membrane and auditory receptors in the locust *Locusta migratoria* L. *J. Exp. Biol.* **198**, 1327–1339 (1995).
- Strausfeld, N. J. & Miller, T. A. *Neuroanatomical Techniques* (Springer, New York, 1980).
- Rheinlaender, J. & Römer, H. Bilateral coding of sound direction in the CNS of the bushcricket *Tettigonia viridissima* L. (Orthoptera, Tettigoniidae). *J. Comp. Physiol. A* **140**, 101–111 (1980).

Acknowledgements. We thank the Alexander family, B.-A. Gereben-Krenn, M. Rieser, H. Schuster, W. van Staaden and R. Wright for assistance; A. Delago for help with experiments; C. Kernbichler, S. Ott, K. Steiner and Wirsam Scientific & Precision Equipment for technical assistance; the KwaZulu/Natal Parks Board for collecting permits; and G. Boyan, J. Fullard, R. Huber and D. Robert for comments. This research was supported by the Austrian Science Foundation.

Correspondence and requests for materials should be addressed to M.J.v.S. (e-mail: m.van-staaden@zoology.kfunigraz.ac.at).

A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated

Karin Jegalian & David C. Page

Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

Mammalian X and Y chromosomes evolved from an autosomal pair: the X retained and the Y gradually lost most ancestral genes^{1,2}. In females, one X chromosome is silenced by X inactivation, a process that is often assumed to have evolved on a broadly regional or chromosomal basis³. Here we propose that genes or clusters common to both the X and Y chromosomes (X–Y genes) evolved independently along a multistep path, eventually acquiring dosage compensation on the X chromosome. Three genes studied here, and other extant genes, appear to be intermediates. *ZFX*, *RPS4X* and *SMCX* were monitored for X inactivation in diverse species by assaying CpG-island methylation, which mirrors X inactivation in many eutherians. *ZFX* evidently escaped X inactivation in proto-eutherians, which also possessed a very similar Y-linked gene; both characteristics were retained in most extant orders, but not in myomorph rodents. For *RPS4X*, escape from X inactivation seems unique to primates. *SMCX* escapes inactivation in primates and myomorphs but not in several other lineages. Thus, X inactivation can evolve independently for each of these genes. We propose that it is an adaptation to the decay of a homologous, Y-linked gene.

By studying differences and similarities among homologous genes in extant species, one can draw inferences about ancestral genes and map points of evolutionary divergence. In this manner, one can explore the coevolution of the X and Y chromosomes and the evolution of epigenetic phenomena such as X inactivation. We first examined the inactivation status of individual X-linked genes in a wide range of mammalian species. Specialized reagents that are analogous to those used to assess X inactivation in humans and mice are unavailable for other species⁴. However, methylation of CpG islands, which exist at the 5' ends of many genes⁵, proved to be a widely applicable alternative. For human and murine X-linked genes, a perfect correlation has been observed between 5' CpG-island methylation and X-inactivation status: transcriptionally silent, X-inactivated alleles are methylated, whereas active alleles are unmethylated⁶. To determine whether this correlation extends to a broad range of eutherians, we examined CpG-island methylation at *ALD*, an X-linked gene known to be X-inactivated in humans^{7,8}. Eighteen species representing nine eutherian orders were tested. In all 18 species, methylation was observed in females, where *ALD* is presumably silenced on the inactive X chromosome; no methylation was observed in males, where the single X chromosome is active (Figs 1 and 2). This suggests that CpG-island methylation accompanies X inactivation in a wide range of eutherians.

We then explored the status, in diverse mammals, of three X-

linked genes that escape X inactivation in humans and that bear CpG islands that are amenable to methylation assays: *ZFX*, *RPS4X* and *SMCX*. We examined the *ZFX* CpG island^{9,10} in 19 eutherian species representing eleven orders. In 15 species, neither female nor male DNA exhibited methylation, indicating that *ZFX* escapes X

inactivation (Fig. 1). Methylation was observed in females of only four species: mouse (where *Zfx* was previously shown to be X-inactivated^{11,12}), rat, hamster and lemming. These species are monophyletic, having evolved from a common ancestor not shared by any of the other 15 eutherian species examined (Fig. 2).

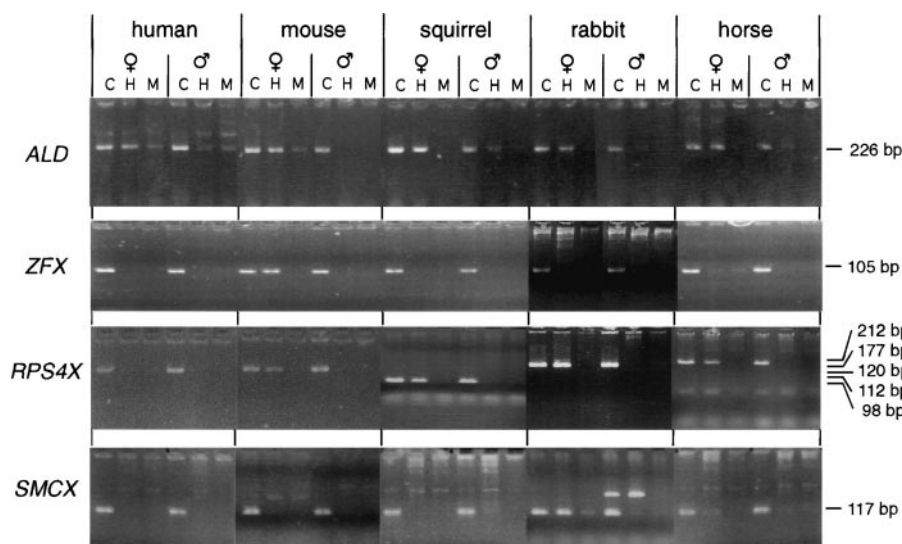


Figure 1 CpG-island methylation studies of four genes in five eutherian species. Female and male genomic DNAs digested with *Hind*III (C, positive 'control') or *Hpa*II (H) or *Msp*I (M, negative control) were used as templates in PCR assays. Expected product sizes (in base pairs) are indicated; *RPS4X* product size differs in

each species. A strong product in the female H lane (comparable to the female C lane, and stronger than the male H lane) indicates that the gene is methylated and X inactivated, whereas a weak or absent product in the female H lane indicates that the gene is not methylated and escapes X inactivation.

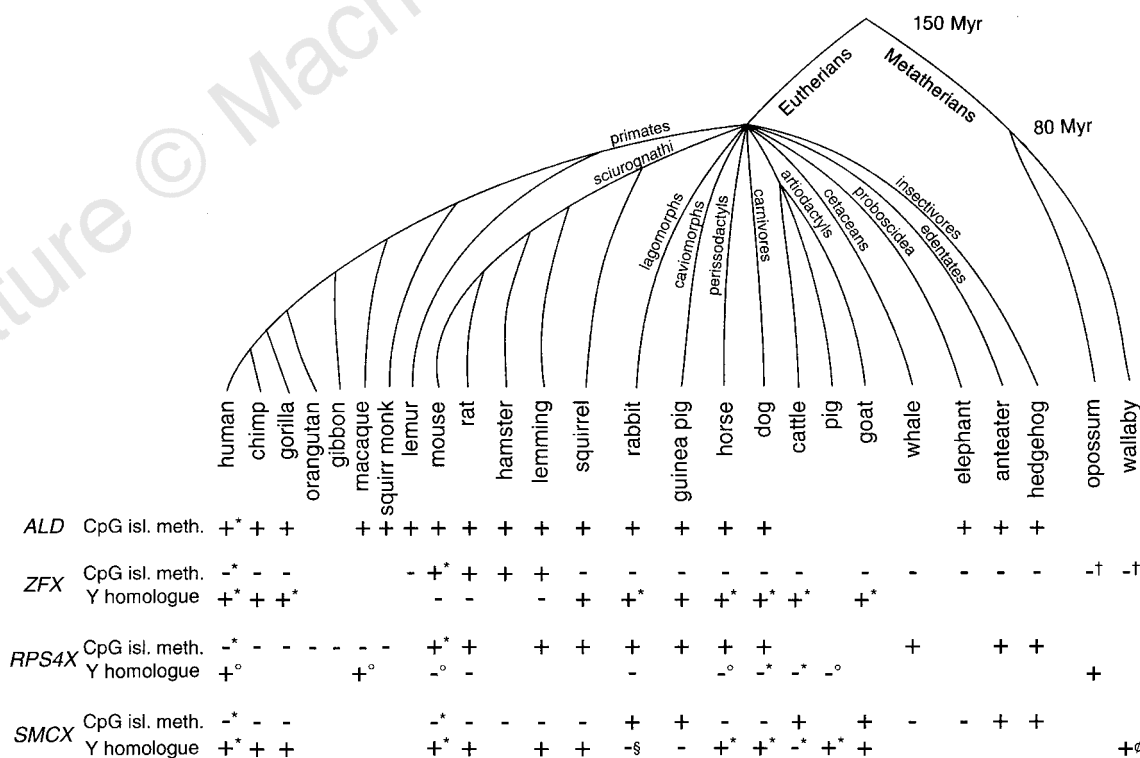


Figure 2 Summary of studies of CpG-island methylation and Y-chromosome homologue in 26 mammalian species, arranged phylogenetically²⁸. Genes were scored '+' for CpG-island methylation if methylation was detected in female but not male DNA, and '-' if methylation was detected in neither sex (Fig. 1). For Y homologues, genes were scored '+' if experiments described in the text and Methods yielded reproducible evidence of highly similar, male-specific counterparts. Asterisk, our results confirmed published findings^{7,9-17,20,22,26}. Other symbols

defined as follows. † CpG island was found to be unmethylated in marsupials, where homologue of human *ZFX* is autosomal²⁹. ° Previously published findings^{11,13,14,20,22}. § Rabbit *SMCY* was detected by Southern blotting (our studies and ref. 20) but appears not to be expressed (see text) * B. Duffy, J. Graves and collaborators, in preparation. Aguinik and collaborators reported evidence of *SMCY* homologue on kangaroo Y chromosome²⁰.

We conclude that *ZFX* escaped X inactivation in the common ancestor of all placental mammals, but became subject to X inactivation in myomorph rodents after the divergence of the sciuriform (squirrel) lineage 40–70 Myr BP.

Like *ZFX*, *RPS4X* is known to escape X inactivation in humans and to undergo X inactivation in mice^{11,13,14}. Given that X inactivation is controlled at the level of the whole chromosome during development, we anticipated that, among diverse mammals, the inactivation status of *RPS4X* might mirror that of *ZFX*. Instead, we found evidence that inactivation of *RPS4X* is much more widespread. Among 11 eutherian orders tested by CpG-island methylation, only primate *RPS4X* (seven species tested) seems to escape X inactivation. In all non-primate species tested (eight eutherian orders), we observed methylation of the *RPS4X* CpG island in female DNA, which probably reflects X inactivation of the gene (Figs 1 and 2).

Finally, we examined *SMCX*, an X-linked gene known to escape inactivation in both mice and humans^{15–17}. Because mice appear to inactivate the X chromosome more completely than do other eutherians^{11,12,14}, we expected that genes escaping inactivation on the mouse X chromosome would do the same in all eutherians. Instead, results from CpG-island methylation studies suggested that *SMCX* is X-inactivated in 5 of 11 eutherian orders (6 of 18 species) tested (Figs 1 and 2).

Our studies yielded several general conclusions. For each of the genes studied—*ZFX*, *RPS4X* and *SMCX*—methylation status was concordant among the four myomorph rodent species tested, and among the seven primate species tested. If CpG-island methylation is a faithful marker of X inactivation, a change in the inactivation behaviour of a given X-linked gene must be only an occasional event during evolution. A single evolutionary transition could account for the X inactivation status of *ZFX* among diverse eutherian species, and no more than a few events could account for the status of either *RPS4X* or *SMCX*. Nonetheless, within a species the methylation status of any one gene—*ZFX*, *RPS4X* or *SMCX*—poorly predicts the other genes' behaviour. Thus, these evolutionary transitions in X inactivation affected individual genes or gene clusters, not whole chromosomes.

Were these occasional changes in X-inactivation status accompanied by the decay of homologous Y-linked genes^{18,19}? If so, the Y-linked homologue should exhibit evidence of functional decline in lineages where the X-linked gene has become subject to X inactivation. We found this for all three X-linked genes under study, each known to have a functional homologue on the human Y chromosome^{9,13,20}. For *ZFX*, Southern blotting revealed a highly similar, male-specific homologue, *ZFY*, in diverse eutherians where *ZFX* is unmethylated (Figs 2 and 3). Only myomorph rodents, where *ZFX* undergoes X inactivation, showed no evidence of highly conserved *ZFY* genes. Based on previous studies in mice, it seems likely that the tested myomorphs possess *ZFY* genes; their nucleotide sequences may differ substantially from those of non-myomorphs because of circumscribed expression and function, and hence relaxed selective pressures²¹. We infer that: (1) a *ZFY* gene, which closely resembled that in humans, existed in proto-eutherians, where *ZFX* escaped X inactivation; (2) these primitive characteristics were retained in most extant eutherian orders, including sciuriforms (squirrels; Figs 2 and 3); and (3) myomorph *ZFY* function became restricted and *ZFX* became subject to X inactivation after the divergence of sciuriforms.

With *RPS4X*, we and other investigators^{11,14,22} searched for Y-chromosomal homologues in diverse eutherians by various means. A homologous Y gene, *RPS4Y*, was identified only in primates, the sole eutherian order in which *RPS4X* is known to escape X inactivation (Fig. 2). These findings are of particular interest given present evidence that *RPS4Y* and *RPS4X* genes coexisted in the common ancestor of eutherians and metatherians (marsupials) 150 Myr BP: in opossum, a marsupial, we identified full-length

cDNA clones from two different *RPS4* genes, one mapping to the X and the other to the Y chromosome, as in humans (K.J. and D.C.P., unpublished results). Thus it appears that proto-eutherians possessed an *RPS4Y* gene. It has been maintained in modern primates, where *RPS4X* escapes X inactivation. *RPS4Y* was lost in other eutherian lineages, where *RPS4X* is now subject to X inactivation.

Like *RPS4X*, homologues of *SMCX* are found on human and marsupial Y (and X) chromosomes, implying that proto-eutherians possessed an *SMCY* gene (ref. 20 and B. Duffy, J. Graves and collaborators, in preparation). Previous and present Southern blotting studies suggest that *SMCY* has been lost, or diverged greatly, in cattle and guinea pigs—both species in which *SMCX* seems to be subject to X inactivation. However, in rabbits and goats, male-specific homologues are detectable by Southern blotting, although *SMCX* is apparently subject to X inactivation (Fig. 2; see also ref. 20). In rabbits, we used screening of cDNA libraries to recover 13 *SMCX* but no *SMCY* cDNA clones. This contrasts with the balanced expression of *SMCX* and *SMCY* in humans and mice^{15,20}. Thus, at least in rabbits, *SMCX* is X inactivated and little or no *SMCY* is expressed.

Our findings based on three X–Y gene pairs suggest that differentiation of X and Y chromosomes occurred on a gene-by-gene or cluster-by-cluster basis, and has been an ongoing process in the eutherian lineages that radiated some 80 Myr BP. Extant eutherian sex chromosomes are only partly differentiated; a subset of genes has retained primitive or intermediate characteristics. We speculate that many extant genes represent intermediates on a general pathway by which eutherian X–Y genes or clusters evolved from autosomal genes (Fig. 4). Autosomal genes were funnelled into the pathway in two ways: by their mere presence on emergent sex chromosomes or by additions to those sex chromosomes through translocation of autosomal material². This was followed by suppression of X–Y recombination. These steps occurred at the chromosomal or regional level, and gave rise to functionally equivalent X-linked (but not X-inactivated) and Y-linked genes. We postulate that subsequent steps occurred on a gene-by-gene or cluster-by-cluster basis. Three processes interacted, perhaps in coupled increments, resulting in an X-linked, X-inactivated gene with complete loss of the Y gene. This

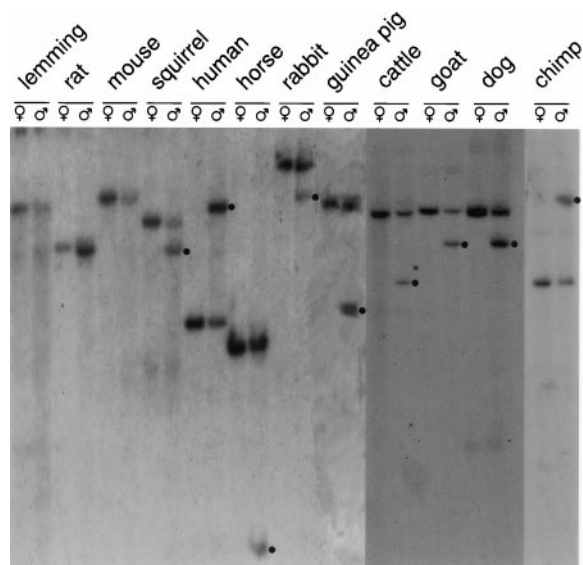


Figure 3 Southern blotting reveals closely related, Y-chromosomal homologue of *ZFX* in many eutherians, but not myomorph rodents. A 395-bp *Bss*III fragment from human *ZFY* CpG island¹⁰ was hybridized to *Eco*RI-digested genomic DNAs (5 µg per lane; female rat lane is underloaded). A black dot is placed immediately to right of each male-specific fragment observed.

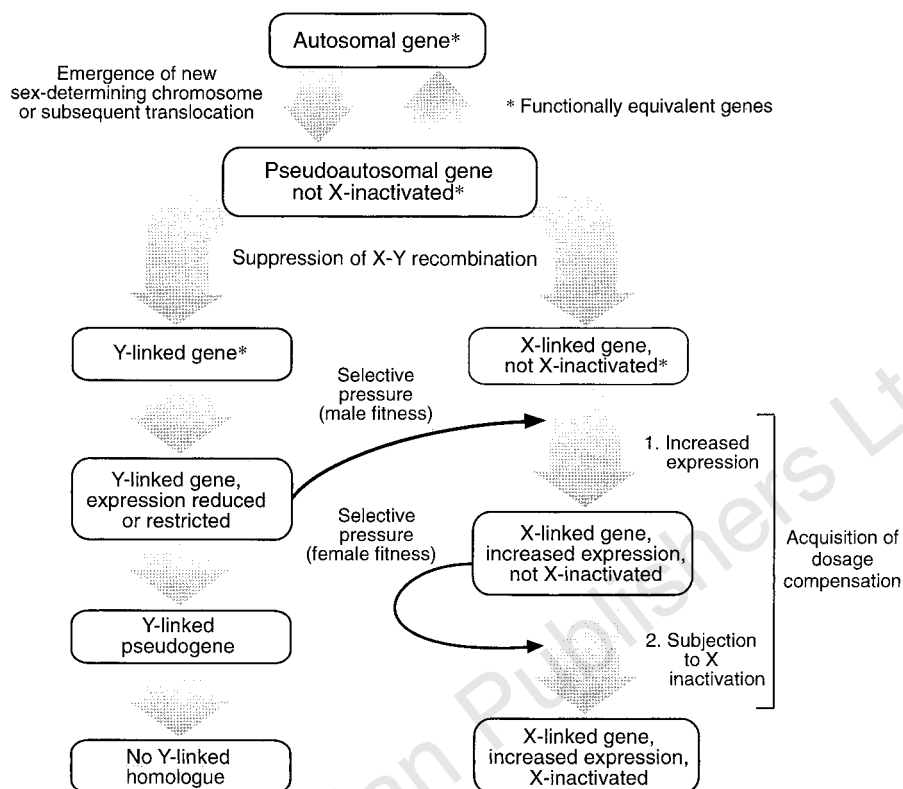


Figure 4 A proposed pathway for X-Y gene evolution in mammals. This could account for most X-linked and X-Y homologous genes in extant mammals, many of these genes existing at intermediate steps in the pathway. We are aware of one

gene that cannot be accounted for by this pathway: the human pseudoautosomal gene *SYBL1*, which is X inactivated and is transcriptionally silenced on the Y chromosome³⁰.

began with Y-gene decay, followed by upregulation of X-linked gene expression, and concluded with X inactivation. Y-gene decay initially may have taken the form of reduced or restricted expression, as in the case of mouse *Zfy* or rabbit *SMCY*. Expression of the X-linked gene or cluster increased as an adaptation to reduced or restricted expression of its Y-linked homologue. This compensated for loss of Y function and restored optimal expression levels in males^{18,23,24}. According to our model, X inactivation was a counter-response, restoring optimal expression levels in females.

Our model provides a more comprehensive explanation than the traditional view, which emphasized that X inactivation was rapidly implemented during evolution on a broadly regional or chromosome-wide basis^{1,3}. Our model conflicts with the hypothesis that rapid evolutionary spreading of X inactivation preceded Y-gene decay and drove its initial steps². This 'X-driven' hypothesis predicts that inactivated X-linked genes with functionally comparable Y-linked homologues should exist as evolutionary intermediates; to our knowledge, such genes have not been found in any mammal. To the contrary, recent studies suggest that numerous human X-linked genes escape X inactivation but have no detectable Y counterparts⁸. The 'Y-driven' pathway of X-Y gene evolution that we propose readily accounts for such genes as intermediates. Our model is not at odds with the view that an X-inactivation signal spread rapidly through large chromosomal regions. However, we speculate that this signal initially had no effect on the expression of many X-linked genes or clusters, perhaps because the required gene- or domain-specific regulatory elements²⁵ did not exist. Local and longer-range regulatory elements subsequently evolved, bringing with them X inactivation on a gene-by-gene or cluster-by-cluster basis. However, our comparative X-Y data suggest that the evolution of X-linked gene expression was more than the random playing out of such regulatory influences. The path was evolutionarily driven and constrained, through natural selection, by coupled changes on the Y chromosome. □

Methods

CpG island methylation. Methylation was assayed by polymerase chain reaction (PCR) using *HindIII*, *HpaII*- or *MspI*-digested genomic DNAs as templates. DNA was prepared from human lymphoblastoid cell lines; BALB/c mouse spleen or liver; liver from Sprague-Dawley rat, Syrian gold hamster, lemming (*Myopus schisticolor*), pilot whale, giant anteater, hedgehog (*Erinaceus europaeus*), wallaby (*Macropus eugenii*) and opossum (*Monodelphis domestica*); squirrel (*Sciurus carolinensis*) kidney; Hartley guinea pig kidney or liver; and blood of all other species tested. For PCR, 50 ng template DNA in 20 µl of 2.5 mM MgCl₂, 5 mM NH₄Cl, 12.5 mM NaCl, 50 mM KCl, 12.5 mM Tris pH 8.2 and 1 µM of each primer, was heated to 100 °C for 5 min before adding dNTPs (to 125 µM each) and 2 U Taq polymerase. Thirty cycles of 1 min at 94 °C, 1 min at 62 °C, and 1 min at 72 °C were followed by extension for 2 min at 72 °C, unless otherwise specified. PCR products were subjected to electrophoresis in 4% NuSieve agarose (FMC Corp.), 90 mM Tris-borate pH 8.3, 2 mM EDTA, 0.5 mg l⁻¹ ethidium bromide and visualized with UV light. All assays were repeated multiple times, in most cases on multiple male and female samples from the same species, with concordant results. Some X-linked genes may escape X inactivation in a subset of cells or developmental stages^{4,8}, a subtlety that these studies of adult tissue methylation would not detect.

For *ALD*, PCR primers were chosen from CpG-island sequences found to be conserved between human and mouse: GTGACATGCCGGTGCTCTCCA and CGCTGCAGGAATACCCGGTTCAT amplify a 226 base pair (bp) product that includes six CCGG (*HpaII/MspI*) sites in humans and three CCGG sites in mice. The PCR annealing temperature was 60 °C, and 0.5 µl DMSO was added to each reaction. For *ZFX*, we previously described PCR primers that span a highly conserved 105-bp region containing two CCGG sites¹⁰. For *RPS4X*, a unique set of primers was required for nearly every species. First, CpG-island sequences for PCR-amplified from genomic DNAs using primers CCTA(G/A)CGCAGCCATGGTAAG (overlapping the exon 1/intron 1 boundary) and TCTTGGGACCACGAGCCT (in exon 2). (For dog, the 5' primer was GCCTCGCGCAGCCATGGT.) Sequencing of these PCR products confirmed the presence of a CpG island in intron 1 in each species and enabled us to select

primers with which to assay *RPS4X* methylation in particular species; primer and underlying sequences have been deposited at GenBank: human, accession no. G36429; chimp and gorilla, G36430 and G36431; rabbit, G36432; guinea pig, G36433; mouse, G36434; rat, G36435; lemming, G36436; squirrel, G36437; dog, G36438; anteater, G36439; hedgehog, G36440; whale, G36441; and horse, G36442. For *SMCX*, PCR primers were chosen from CpG-island sequences found to be conserved between human and mouse: CCTCGGGGCCACCATG-GAG and CTGATTTTCGCGATGTAGCC amplify a 117-bp product that includes three CCGG sites in humans and two in mice. We selected these conserved *SMCX* primers after sequencing 5' portions of the mouse transcript (GenBank AF0398940; obtained by 5' RACE cloning) and comparing these with previously published 5' human *SMCX* sequences²⁰.

Y-chromosome homologues. We searched for Y-specific homologues of *ZFX/ZFY* and *SMCX/SMCY* in mammalian species by Southern blotting of *EcoRI*-digested male and female genomic DNAs. For *ZFY*, we used two hybridization probes in separate experiments, with entirely concordant results (Figs 2 and 3): (1) a 395-bp genomic *Bss*HII fragment¹⁰ from the 5' CpG island of human *ZFY*, and (2) pDPI007, a 1.3-kb genomic fragment containing the zinc-finger exon²⁶ of human *ZFY*. Probes were labelled with ³²P by random-primer synthesis and hybridized overnight to Southern blots at 67 °C (65 °C for pDPI007) in 1 mM EDTA, 0.5 M NaPO₄ pH 7.2 and 7% sodium dodecyl sulphate (SDS). Blots were then washed three times for 20 min each at 62 °C in 0.1 × SSC (1 × SSC = 0.15 M NaCl, 15 mM Na citrate pH 7.4), 0.1% SDS and exposed at -80 °C with X-ray film backed with an intensifying screen for one day. The *SMCY* hybridization probe (pCM4) and conditions were described previously²⁰.

We also searched for Y-chromosome homologues of *RPS4X/RPS4Y* and *SMCX/SMCY* in particular species by cDNA selection, or by screening cDNA libraries. In cDNA selection²⁷, human *RPS4Y* coding sequence (as selector) was hybridized at 55 °C to cDNA libraries (Clontech) prepared from adult male rat, rabbit, dog or cattle liver. Selection products were cloned into plasmid vectors and sequenced. cDNA libraries prepared from adult male dog liver (Clontech) and adult male opossum spleen (Stratagene) were screened with the entire human *RPS4Y* coding sequence as probe, at low stringency (overnight hybridization at 58 °C in 1 mM EDTA, 0.5 M NaPO₄ pH 7.2, 7% SDS; subsequent washing three times for 20 min each at 50 °C in 1 × SSC, 0.1% SDS). Once dog and opossum *RPS4X* clones were identified, by sequencing, these were used as probes for high-stringency rescreening of their respective libraries (hybridization at 65 °C and washes at 65 °C in 0.1 × SSC, 0.1% SDS). We anticipated that *RPS4Y* clones, if present, would be detected in the low-stringency screen but not in the high-stringency screen. In this manner, we identified opossum *RPS4Y* (and *RPS4X*) cDNA clones, confirmed by sequencing (GenBank AF051137 and AF051136, respectively) and mapping studies (K.J. and D.C.P., unpublished results; complete description will be published elsewhere), but in dog we detected only *RPS4X* clones. Similarly, we screened a cDNA library (Clontech) prepared from adult male rabbit liver at low stringency using a 370-bp mouse *Smcx* cDNA fragment (prepared from clone pCM4 (ref. 20) by PCR using primers CCTTCCAAGTTCAACAGTTATGG and CATACGTATGACTCAATAAAGTGGG), identifying 13 *SMCX* but no *SMCY* clones.

Received 25 March; accepted 9 June 1998.

- Ohno, S. *Sex Chromosomes and Sex-linked Genes* (Springer, Berlin, 1967).
- Graves, J. A. & Schmidt, M. M. Mammalian sex chromosomes: design or accident? *Curr. Opin. Genet. Dev.* **2**, 890–901 (1992).
- Lyon, M. F. Evolution of X-chromosome inactivation in mammals. *Nature* **250**, 651–653 (1974).
- Disteche, C. M. The great escape. *Am. J. Hum. Genet.* **60**, 1312–1315 (1997).
- Cross, S. H. & Bird, A. P. CpG islands and genes. *Curr. Opin. Genet. Dev.* **5**, 309–314 (1995).
- Tribioli, C. *et al.* Methylation and sequence analysis around *EagI* sites: identification of 28 new CpG islands in Xq24-Xq28. *Nucl. Acids Res.* **20**, 727–733 (1992).
- Migeon, B. R. *et al.* Adrenoleukodystrophy: evidence for X linkage, inactivation, and selection favoring the mutant allele in heterozygous cells. *Proc. Natl Acad. Sci. USA* **78**, 5066–5070 (1981).
- Brown, C. J., Carrel, L. & Willard, H. F. Expression of genes from the human active and inactive X chromosomes. *Am. J. Hum. Genet.* **60**, 1333–1343 (1997).
- Schneider-Gädick, A., Beer-Romero, P., Brown, L. G., Nussbaum, R. & Page, D. C. *ZFX* has a gene structure similar to *ZFY*, the putative human sex determinant, and escapes X inactivation. *Cell* **57**, 1247–1258 (1989).
- Luoh, S. W. *et al.* CpG islands in human *ZFX* and *ZFY* and mouse *Zfx* genes: sequence similarities and methylation differences. *Genomics* **29**, 353–363 (1995).
- Ashworth, A., Rastan, S., Lovell-Badge, R. & Kay, G. X-chromosome inactivation may explain the difference in viability of XO humans and mice. *Nature* **351**, 406–408 (1991).
- Adler, D. A., Bressler, S. L., Chapman, V. M., Page, D. C. & Disteche, C. M. Inactivation of the *Zfx* gene on the mouse X chromosome. *Proc. Natl Acad. Sci. USA* **88**, 4592–4595 (1991).
- Fisher, E. M. *et al.* Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* **63**, 1205–1218 (1990).

- Zinn, A. R. *et al.* Inactivation of the *Rps4* gene on the mouse X chromosome. *Genomics* **11**, 1097–1101 (1991).
- Aguinik, A. I. *et al.* A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. *Hum. Mol. Genet.* **3**, 879–884 (1994).
- Wu, J. *et al.* Isolation and characterization of *XEL169*, a novel human gene that escapes X-inactivation. *Hum. Mol. Genet.* **3**, 153–160 (1994).
- Wu, J. *et al.* The murine *Xel169* gene escapes X-inactivation like its human homologue. *Nature Genet.* **7**, 491–496 (1994).
- Charlesworth, B. The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* **6**, 149–162 (1996).
- Rice, W. R. Evolution of the Y sex chromosome in animals. *BioScience* **46**, 331–343 (1996).
- Aguinik, A. I., Mitchell, M. J., Lerner, J. L., Woods, D. R. & Bishop, C. E. A mouse Y chromosome gene encoded by a region essential for spermatogenesis and expression of male-specific minor histocompatibility antigens. *Hum. Mol. Genet.* **3**, 873–878 (1994).
- Mahaffey, C. L. *et al.* Intron/exon structure confirms that mouse *Zfy1* and *Zfy2* are members of the *ZFY* gene family. *Genomics* **41**, 123–127 (1997).
- Omoe, K. & Endo, A. Relationship between the monosomy X phenotype and Y-linked ribosomal protein S4 (*Rps4*) in several species of mammals: a molecular evolutionary analysis of *Rps4* homologs. *Genomics* **31**, 44–50 (1996).
- Muller, H. J. Evidence of the precision of genetic adaptation. *Harvey Lect.* **43**, 165–229 (1950).
- Adler, D. A. *et al.* Evidence of evolutionary up-regulation of the single active X chromosome in mammals based on *Cle4* expression levels in *Mus spretus* and *Mus musculus*. *Proc. Natl Acad. Sci. USA* **94**, 9244–9248 (1997).
- Goldman, M. A. The chromatin domain as a unit of gene regulation. *BioEssays* **9**, 50–55 (1988).
- Page, D. C. *et al.* The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **51**, 1091–1104 (1987).
- Lovett, M., Kere, J. & Hinton, L. M. Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl Acad. Sci. USA* **88**, 9628–9632 (1991).
- Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356**, 121–125 (1992).
- Sinclair, A. H. *et al.* Sequences homologous to *ZFY*, a candidate human sex-determining gene, are autosomal in marsupials. *Nature* **336**, 780–783 (1988).
- D'Esposito, M. *et al.* A synaptobrevin-like gene in the Xq28 pseudoautosomal region undergoes X inactivation. *Nature Genet.* **13**, 227–229 (1996).

Acknowledgements. We thank the New England Aquarium, the San Diego Zoo, the Duke Primate Center, K. Fredga and K. Campbell for help in obtaining animal specimens; A. Aguinik for a mouse *Smcx* cDNA clone; D. Graur and B. Lahn for helpful discussions; J. Graves for permission to cite unpublished work; J. Cook-Chrysos for graphic art; and D. Bartel, B. Charlesworth, A. Chess, C. Disteche, G. Fink, J. Graves, R. Jaenisch, T. Kawaguchi, M. L. Pardue, D. Schlessinger, C. Sun, H. Willard and A. Zinn for comments on the manuscript. This work was supported by the National Institutes of Health and the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to D.C.P. (e-mail: dcpag@wi.mit.edu).

Signal-dependent noise determines motor planning

Christopher M. Harris* & Daniel M. Wolpert†

* Department of Ophthalmology and Visual Sciences Unit, Great Ormond Street Hospital for Children NHS Trust, and Institute of Child Health, University College London, London WC1N 3JH, UK

† Sobell Department of Neurophysiology, Institute of Neurology, Queen Square, University College London, London WC1N 3BG, UK

When we make saccadic eye movements or goal-directed arm movements, there is an infinite number of possible trajectories that the eye or arm could take to reach the target^{1,2}. However, humans show highly stereotyped trajectories in which velocity profiles of both the eye and hand are smooth and symmetric for brief movements^{3,4}. Here we present a unifying theory of eye and arm movements based on the single physiological assumption that the neural control signals are corrupted by noise whose variance increases with the size of the control signal. We propose that in the presence of such signal-dependent noise, the shape of a trajectory is selected to minimize the variance of the final eye or arm position. This minimum-variance theory accurately predicts the trajectories of both saccades and arm movements and the speed-accuracy trade-off described by Fitt's law⁵. These profiles are robust to changes in the dynamics of the eye or arm, as found empirically^{6,7}. Moreover, the relation between path curvature and hand velocity during drawing movements reproduces the empirical 'two-thirds power law'^{8,9}. This theory provides a simple and powerful unifying perspective for both eye and arm movement control.

The trajectories of eye and arm movements (that is, the change in position and velocity over time) are not inevitable consequences of

the mechanical properties of muscles, but reflect an orchestrated pattern of neural activation by motor and pre-motor neurons. For saccadic eye movements, it has been proposed that trajectories are selected to minimize the time to reach the target¹⁰. For a linear system, the minimum-time requirement results in 'bang-bang' control, where the control signal is instantaneously switched between its maximum positive and negative values to accelerate and decelerate the eye. However, it is difficult to generate the observed symmetrical saccadic velocity profiles with bang-bang control signals^{11,12}. For arm movements, it has been suggested that trajectories are selected to optimize a cost that is integrated over the movement, such as jerk (rate of change of acceleration)^{13,14} or torque change¹⁵. However, there has been no principled explanation why the central nervous system should have evolved to optimize such quantities, other than that these models predict smooth trajectories. Indeed, the advantage of smoothness of movement still remains unexplained. Furthermore, how the central nervous system could estimate complex quantities, such as jerk or torque change, and then integrate them over the duration of a trajectory is also unknown.

We propose that minimizing the variance of the eye or arm's

position, in the presence of biological noise, is the underlying determinant of trajectory planning. Noise in the final neural control signal (that is, noise in the firing of motor neurons) will cause trajectories to deviate from the desired path. These deviations will be accumulated over the duration of a movement, leading to variability in the final position. If the noise were independent of the control signal¹⁶, then the accumulated error would be minimized by making the movement as rapidly as possible, as in bang-bang control for linear systems. However, here we assume that the noise in the neural control signal increases with the mean level of the signal. This is based on the empirical observation that the standard deviation of motor-neuronal firing increases with the mean level, with a coefficient of variation between 10 and 25% (refs 17, 18). This assumption of signal-dependent noise is also consistent with psychophysical observations that the variability of motor errors increases with the magnitude of the movement, as captured by the empirical Fitt's law⁵. In the presence of such signal-dependent noise, moving as rapidly as possible requires large control signals, which would increase the variability in the final position. As the resulting inaccuracy of the movement may lead to task failure or require further corrective movements, moving very fast becomes counterproductive^{19,20}. Accuracy could be improved by having low control signals, but the movement will be slow. Thus, signal-dependent noise inherently imposes a trade-off between movement

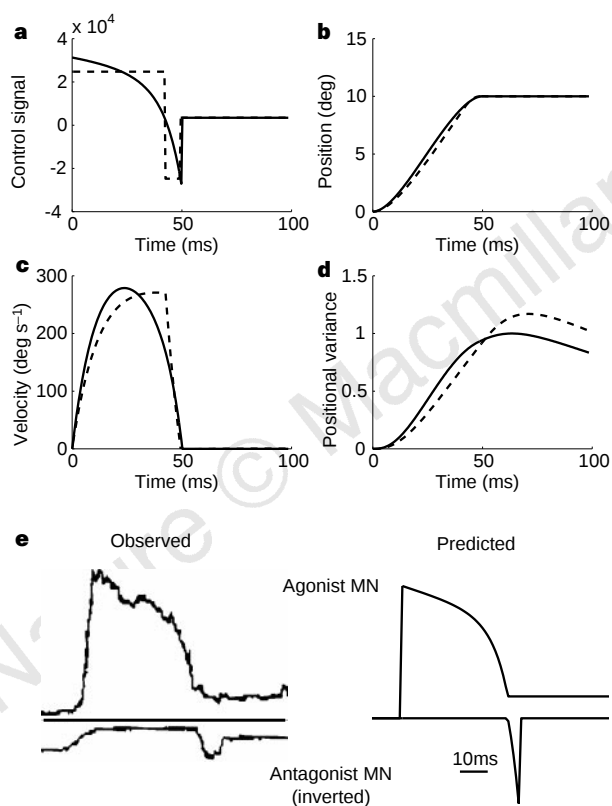


Figure 1 A comparison of bang-bang (dashed line) and minimum-variance (solid line) control for a 10 degree saccade with a duration of 50 ms. **a**, Neural control signals; **b**, average position profiles of trajectories across repeated movements are similar; **c**, velocity profiles show a marked difference in their degree of symmetry; **d**, positional variance (obtained from equation (1) in Methods) which represents the spread of the eye position across repeated trials, about the mean position shown in **b**. Note that the variance continues to evolve after the mean position has become steady. The eye was modelled as a second-order linear system with time constants of 224 and 13 ms. The variances have been scaled so that the peak for the minimum-variance model is unity. **e**, Empirical (from ref. 22, with permission) and predicted motor neuronal firing of agonist and antagonist eye muscles in monkey for a 12 degree saccade. To model the monkey's faster plant, time constants of 150 and 7 ms were used for this simulation. The predicted motoneuronal activity was derived from the control signal by splitting it into positive (agonist) and negative (antagonist) parts (the final tonic level of the antagonist has not been modelled). Note that temporal jitter of firing is likely to broaden and lower the predicted sharp antagonist peak.

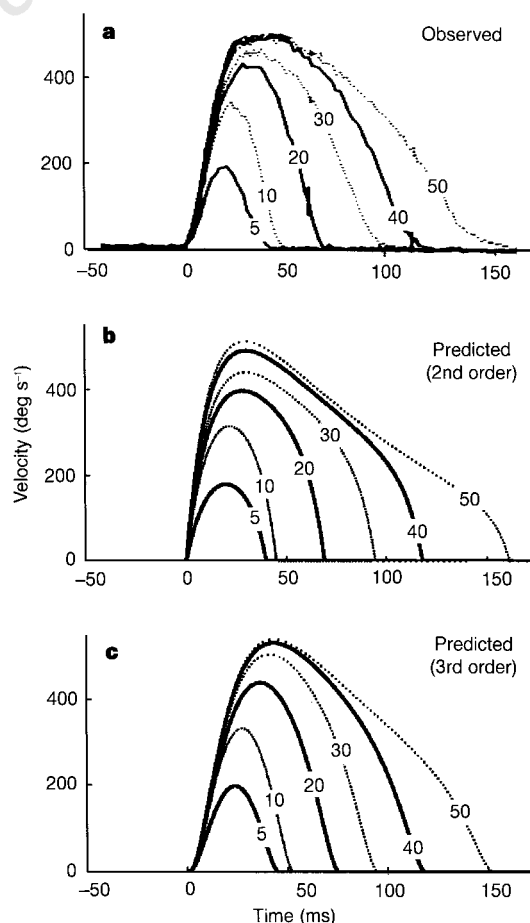


Figure 2 Comparison of empirical and predicted saccade trajectories. **a**, Velocity profiles of actual horizontal saccadic eye movements ranging in amplitude from 5 to 50 degrees (taken from ref. 4, with permission). Saccades below 30 degrees are symmetric, whereas larger saccades are asymmetric with a longer decelerative phase. **b**, Theoretical optimal trajectories for minimizing positional variance with signal-dependent noise for a second-order linear model of the eye with time constants of 224 and 13 ms. **c**, Theoretical optimal trajectories for a third-order linear model of the eye with the additional time constant of 10 ms. The predicted trajectories show the change in symmetry with saccade amplitude.

duration and terminal accuracy. The key point is that for a given amplitude and duration of movement, the final positional variance will depend critically on the actual neural commands and the subsequent velocity profile. We propose that the temporal profile of the neural command is selected so as to minimize the final positional variance for a specified movement duration, or equivalently to minimize the movement duration for a specified final positional variance determined by the task.

We assume that neural commands have signal-dependent noise whose standard deviation increases linearly with the absolute value of the neural control signal. On the basis of this single assumption, we determined the optimal trajectories of the eye and arm that minimized the total positional variance during the immediate post-movement period (see Methods). For saccades, we considered a commonly used second-order linear system²¹, in which the minimum-variance and bang-bang solutions were compared for a 50-ms 10-degree movement (Fig. 1). Both the control strategies reach the target position in the same time (Fig. 1b), but the positional variance in the post-movement period is significantly lower for the minimum-variance solution (Fig. 1d). Although on average the eye is on target, with zero velocity at the end of the movement (Fig. 1b), on any single trajectory there will be a positional error and a non-zero velocity at the end of the movement due to the effect of the signal-dependent noise. Therefore, as seen in Fig. 1d, the positional variance can continue to change after the movement. Thus to minimize deviations from the final position, it is necessary to minimize variance over a post-movement period. The minimum-variance solution results in a smooth symmetric velocity profile, in contrast to the asymmetric profile produced by bang-bang control (Fig. 1c). The model is also in qualitative agreement with observed agonist and antagonist motoneuronal firing patterns²² (Fig. 1e).

Over a range of amplitudes, the minimum-variance solutions (Fig. 2b) also capture the important features of natural saccadic movements (Fig. 2a)—symmetric short movements and asymmetric long movements with an extended deceleration phase⁴. With a third-order model of the eye, the optimal trajectories capture the bell-shaped trajectories observed empirically (Fig. 2c). The optimal trajectories also show a similar rise time for all saccades and a saturation in the peak velocity for saccade amplitudes above ~30 degrees, as is observed experimentally. Although it has been proposed that this saturation arises from limits on the control signal¹⁰, the minimum-variance model predicts this saturation without any such limits. The shapes of trajectories for amplitudes under 20 degrees were insensitive to modest changes in the time constants of the system, whereas these parameters had a much

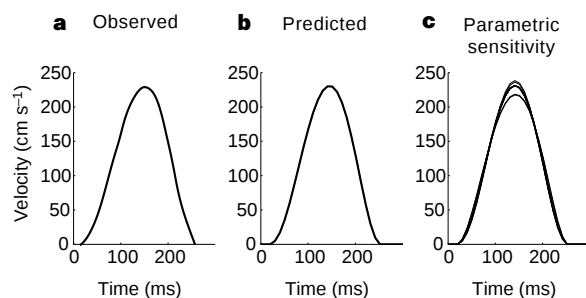


Figure 3 Comparison of empirical and theoretical arm velocity profiles. **a**, Velocity profile of the hand in a typical fast arm movement (taken from ref. 24, with permission). **b**, Theoretical optimal trajectory for minimizing post-movement variance with signal-dependent noise for a second-order skeletal model of a one-dimensional arm with inertia 0.25 kg m^2 and viscosity $0.2 \text{ N ms rad}^{-1}$ driven by a second-order linear muscle with time constants of 30 and 40 ms (parameters taken from ref. 23). **c**, Eight velocity profiles for the model in **b** in which the inertia, viscosity and time constants are individually doubled or halved. The trajectory is essentially invariant to these large changes in the dynamics of the arm.

bigger effect on the slower movements. This is consistent with the general observation that the speed and duration of large saccades have greater variability within and between individuals than those of small saccades. For brief movements, the order of the plant is the major determinant of the optimal trajectory. In the limit, for infinitely brief movements, the highest-order term dominates and the model predicts symmetrical trajectories¹². In Fig. 2, trajectories for second- and third-order models are shown; the trajectories for models of higher order are very similar to the third-order model. For longer movements, the lower-order terms become more important, leading to asymmetrical and complex optimal trajectories.

For arm movements, we first considered a simple one-dimensional fourth-order linear model which captures the muscle dynamics and the arm's skeletal inertia and viscosity²³ (see Methods). The optimal trajectory for a fast movement (Fig. 3b), in which feedback was assumed to play a negligible role, shows the typical bell-shaped velocity profile seen in natural movements²⁴ (Fig. 3a). The profile is again insensitive to changes in the plant, so that even when the inertia and viscosity of the arm or the time constants of the muscle are individually halved or doubled, the optimal profile remains essentially unchanged (Fig. 3c). This is consistent with the observation that when the arm is subject to elastic, viscous or inertial loads, the bell-shaped velocity profile is regained after a period of adaptation^{6,7,25–28}.

For arm movements, the required accuracy varies with the task, as for example in the difference between pointing to someone in a room compared with threading a needle. When pointing at targets, it is empirically observed that movement duration increases with the accuracy demanded by the task. This relationship is captured empirically by Fitt's law, in which the required accuracy is determined by the width of the target (Fig. 4a). For any given movement duration, signal-dependent noise places a lower limit on the final positional variance given by the minimum-variance trajectory. Conversely, given a movement accuracy constraint, specified in terms of final positional variance, there is a minimum duration of movement which can achieve this. We propose that this minimum duration trajectory will be chosen, given the limits of the final endpoint variance imposed by the task. By assuming a linear relationship between the required final endpoint standard deviation

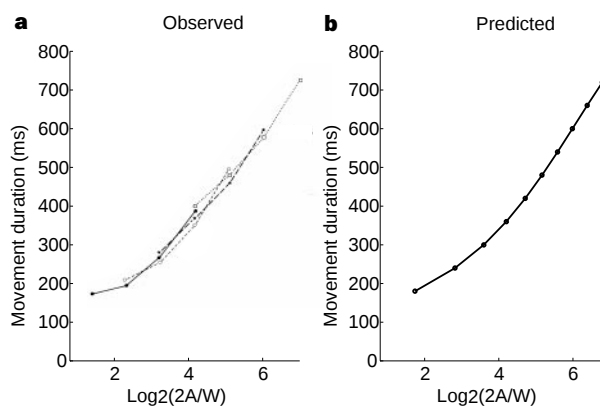


Figure 4 Comparison of empirical and predicted arm movement durations. **a**, Fitt's law relates movement duration (T) to the movement amplitude (A) and target width (W shown as different symbols and lines) according to $T = a + b \log_2(2A/W)$, where a and b are constants. Empirical data from ref. 30 (with permission). This shows the typical increase in movement duration as either the amplitude of movement increases or the target width decreases. **b**, Predicted movement durations for the minimum-variance model in the presence of signal-dependent noise. The movement duration was calculated as the shortest possible, given that the target width places an upper limit on the endpoint positional standard deviation (see Methods). Note the predictions for different widths, W , completely overlap.

and the target width, Fitt's law emerges as a consequence of signal-dependent noise (Fig. 4b).

We next considered a two-joint nonlinear model of the arm moving in the horizontal plane (see Methods). The optimal trajectories for point-to-point movements again showed bell-shaped velocity profiles (Fig. 5d) with gently curved hand paths (Fig. 5b), as observed empirically (Fig. 5a, c)¹⁵. For drawing movements, it has been found that hand velocity (V) decreases as the radius of curvature of the path (R) decreases—this has been formulated as the 'two-thirds power law', $V = KR^{(1-\beta)}$, where K is a constant and $\beta \approx 2/3$ (refs 8, 9). We derived the trajectory that would minimize the positional variance of the hand when repetitively drawing ellipses (see Methods). The optimal trajectory shows the typical slowing down as the curvature increases (Fig. 5e) and reproduces the two-thirds power law, as seen by the regression line in Fig. 5f (slope of 0.32, giving $\beta = 0.68$). This shows that the minimum-variance trajectory predicts the two-thirds power law.

From these analyses, we see that the trajectories of both saccadic eye movements and arm movements can be described as trajectories that minimize post-movement variance in the presence of signal-dependent noise on the control signal. This approach has several important ramifications. Primarily, it provides a biologically plau-

sible theoretical underpinning for both eye and arm movements. In contrast, it is difficult to reconcile observed saccade trajectories with bang-bang control, or to explain the biological relevance of such factors as jerk or torque change in previous models of arm trajectories.

Moreover, there is no need for the central nervous system to construct highly derived signals to estimate the cost to the movement, which is now variance of the final position or the consequences of this inaccuracy, such as the time spent in making corrective movements^{19,20}. Such costs are directly available to the nervous system and the optimal trajectory could be learnt from the experience of repeated movements. Finally, it can be seen that optimal trajectories are inherently smooth. Abrupt changes in the trajectory of the eye or arm require large driving signals which would carry more noise and therefore be suboptimal.

The minimum-variance theory provides a simple, unifying and powerful principle that can be applied to goal-directed movements and implies that signal-dependent noise plays a fundamental role in motor planning. □

Methods

We found the optimal trajectories numerically for both linear models of the eye and arm, and a nonlinear model of a two-joint arm, in the presence of signal-dependent white noise in the control signal. The cost function that was minimized was the positional variance across repeated movements summed over the post-movement period.

Linear models. For the linear models of the eye and arm, we consider a single-input single-output discrete-time system under control with a state-update equation given by $\mathbf{x}_{t+1} = A\mathbf{x}_t + B(u_t + w_t)$, where $\mathbf{x}_t$ is the n -dimensional state at time t , u_t is the neural driving signal at t (note that for arm movements, u_t is the neural command signal that activates muscles and is not torque). A is a fixed $n \times n$ matrix and B is a $n \times 1$ vector describing the dynamics of the system; w_t represents white noise on the driving signal, with zero mean and variance ku_t^2 , which increases with the magnitude of the control signal, u_t , and represents the signal-dependent noise. By iterating the state-update equation, the distribution of the state at time t can be shown to have a mean

$$E[\mathbf{x}_t] = A^t \mathbf{x}_0 + \sum_{i=0}^{t-1} A^{t-1-i} B u_i$$

with covariance

$$\text{Cov}[\mathbf{x}_t] = k \sum_{i=0}^{t-1} (A^{t-1-i} B)(A^{t-1-i} B)^T u_i^2 \quad (1)$$

The variance of the position at time t , V_t , is given by the appropriate element of the diagonal of $\text{Cov}[\mathbf{x}_t]$. We wish to find the driving signal, $\mathbf{u} = [u_0, u_1, u_2, \dots, u_{T+R}]^T$, that reaches the desired position at time step T (the movement time) and maintains it for R steps (the post-movement time) and which minimizes the summed positional variance during this post-movement period, $\sum_{t=T+1}^{T+R} V_t$ (the cost). This can be formulated as a quadratic programming problem which was solved using Matlab.

For the second-order linear model of the eye, the time constants were 224 and 13 ms (ref. 21). Movement durations were taken from ref. 4. Simulations were performed with a time step of 1 ms and a post-movement fixation period of 50 ms. For the third-order model of the eye, an additional time constant of 10 ms was included. There were negligible changes in the optimal trajectories for post-movement times $R > 20$ ms.

For the one-dimensional arm model, we used the combination of a second-order linear model of muscle and a second-order linear model of the arm's skeletal system which included inertia and viscosity. The time constants of the muscle were taken as 40 and 30 ms and the inertia of the arm was 0.25 kg m^2 and the viscosity was $0.2 \text{ N ms rad}^{-1}$ (ref. 23). Simulations were performed with a time step of 10 ms and a post-movement period of 500 ms. There were negligible changes in the optimal trajectories for post-movement times $R > 200$ ms.

To simulate Fitt's law, we assumed that subjects are required to place any part of their finger of width w (taken as 6 mm) within the target of width W with a fixed probability (the success rate). The required accuracy of the movement,

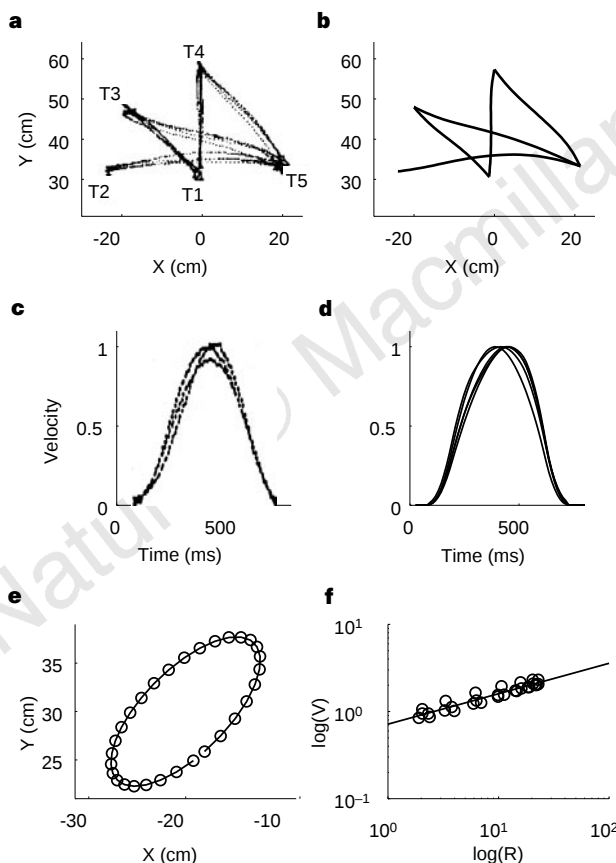


Figure 5 Comparison of empirical and predicted trajectories for a two-joint arm. **a**, Observed hand paths for a set of point-to-point movements (from ref. 15, with permission), and **b**, theoretical optimal hand paths for the same set of movements. The coordinates are centred on the shoulder joint and X and Y directions represent the transverse and sagittal axes respectively. **c**, Observed velocity profiles from T1 to T3 movement in **a**. Note all other movements in **a** showed similar velocity profiles. **d**, Velocity profiles of all the optimal movements shown in **b**, normalized to have a maximum velocity of 1. **e**, Optimal trajectory for drawing an ellipse, the points are equally spaced in time. Note the slowing down at points of high curvature. **f**, Plot of log tangential velocity (V) against log radius of curvature (R). The two-thirds power law predicts $\log(V) = \log(K) + (1 - \beta)\log(R)$. Linear regression gives a slope of 0.32 ($r^2 = 0.85$) and hence the value of $\beta = 0.68$.

specified as the desired final positional standard deviation, was taken as $(W + w)/r$, where r was taken as 1.96 to achieve a 95% success rate. The movement duration was calculated as the shortest possible time that could achieve this accuracy constraint, given that the signal-dependent noise on the entire motor-neuronal pool had a 1% coefficient of variation.

Nonlinear model. For the nonlinear two-jointed planar arm, we used two linear second-order muscles, as described above, acting on the shoulder and elbow joint of a two-link arm moving in the horizontal plane (arm parameters from ref. 29). The trajectories were parametrized as cubic splines with the knots evenly spaced in time. For the point-to-point movements, 7 cartesian (x, y) knots were used with the first and last points fixed at the start and target locations with zero velocity. 500 movements (650-ms duration, sampled at 10 ms) were simulated with signal-dependent noise to determine the trajectory that minimizes the post-movement variance. The optimal trajectory was found using the simplex algorithm to adjust the knot locations.

For ellipse-drawing movements (duration 600 ms, sampled at 20 ms), the knots represented the proportion of the distance travelled around the ellipse as a function of time. Seven knots were used with the first knot at zero and the last at one. This spline determined the velocity profile of the movement which was confined to an elliptic path. The simplex algorithm was used to find the optimal trajectory.

Received 25 March; accepted 8 June 1998.

- Bernstein, N. *The Coordination and Regulation of Movements* (Pergamon, London, 1967).
- Wolpert, D. M. Computational approaches to motor control. *Trends Cogn. Sci.* **1**, 209–216 (1997).
- Morasso, P. Spatial control of arm movements. *Exp. Brain Res.* **42**, 223–227 (1981).
- Collewijn, H., Erkelens, C. J. & Steinman, R. M. Binocular coordination of human horizontal saccadic eye-movements. *J. Physiol.* **404**, 157–182 (1988).
- Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movements. *J. Exp. Psychol.* **47**, 381–391 (1954).
- Shadmehr, R. & Mussa-Ivaldi, F. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
- Brashers-Krug, T., Shadmehr, R. & Bizzi, E. Consolidation in human motor memory. *Nature* **382**, 252–255 (1996).
- Lacquaniti, F., Terzuolo, C. A. & Viviani, P. The law relating kinematic and figural aspects of drawing movements. *Acta Psychologica* **54**, 115–130 (1983).
- Viviani, P. & Schneider, R. A developmental study of the relationship between geometry and kinematics in drawing movements. *J. Exp. Psychol. HPP* **17**, 198–218 (1991).
- Enderle, J. D. & Wolfe, J. W. Time-optimal control of saccadic eye-movements. *IEEE Trans. Biomed. Eng.* **34**, 43–55 (1987).
- Harris, C. M., Wallman, J. & Scudder, C. A. Fourier analysis of saccades in monkeys and humans. *J. Neurophysiol.* **63**, 877–886 (1990).
- Harris, C. M. On the optimal control of behaviour: A stochastic perspective. *J. Neurosci. Meth.* **83**, 73–88 (1998).
- Hogan, N. An organizing principle for a class of voluntary movements. *J. Neurosci.* **4**, 2745–2754 (1984).
- Flash, T. & Hogan, N. The co-ordination of arm movements: An experimentally confirmed mathematical model. *J. Neurosci.* **5**, 1688–1703 (1985).
- Uno, Y., Kawato, M. & Suzuki, R. Formation and control of optimal trajectories in human multijoint arm movements: Minimum torque-change model. *Biol. Cybern.* **61**, 89–101 (1989).
- Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. An internal model for sensorimotor integration. *Science* **269**, 1880–1882 (1995).
- Clamann, P. H. Statistical analysis of motor unit firing patterns in human skeletal muscle. *Biophysics J.* **9**, 1233–1251 (1969).
- Matthews, P. B. C. Relationship of firing intervals of human motor units to the trajectory of post-spike after-hyperpolarization and synaptic noise. *J. Physiol.* **492**, 597–628 (1996).
- Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E. & Smith, J. E. K. Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychol. Rev.* **98**, 340–370 (1988).
- Harris, C. M. Does saccadic under-shoot minimize saccadic flight-time? A Monte Carlo study. *Vision Res.* **35**, 691–701 (1995).
- Robinson, D. A., Gordon, J. L. & Gordon, S. E. A model of the smooth pursuit eye movement system. *Biol. Cybern.* **55**, 43–57 (1986).
- van Gisbergen, J. A. M., Robinson, D. A. & Gielen, C. C. A. M. A quantitative analysis of generation of saccadic eye movements by burst neurons. *J. Neurophysiol.* **45**, 417–442 (1981).
- van der Helm, F. C. T. & Rozendaal, L. A. in *Biomechanics and Neural Control of Movement* (eds Winters, J. M. & Crago, P. E.) (Springer, New York, in the press).
- Kelso, J. A. S., Southard, D. L. & Goodman, D. On the nature of human interlimb coordination. *Science* **203**, 1029–1031 (1979).
- Lackner, J. R. & DiZio, P. Rapid adaptation to Coriolis force perturbations of arm trajectory. *J. Neurophysiol.* **72**, 299–313 (1994).
- Sainburg, R. L. & Ghez, C. Limitations in the learning and generalization of multi-joint dynamics. *Soc. Neurosci. Abstr.* **21**, 686 (1995).
- Flash, T. & Gurevich, I. in *Self-Organization, Computational Maps and Motor Control* (eds Morasso, P. G. & Sanguineti, V.) 423–481 (Elsevier, Amsterdam, 1997).
- Goodbody, S. J. & Wolpert, D. M. Temporal and amplitude generalization in motor learning. *J. Neurophysiol.* **79**, 1825–1838 (1998).
- Kawato, M. in *Attention and Performance, XVI* (eds Inui, T. & McClelland, J.) 335–367 (MIT Press, Cambridge, MA, 1996).
- Jeannerod, M. *The Neural and Behavioural Organization of Goal-directed Movements* (OUP Psychol. Ser. No. 5, Oxford, 1988).

Acknowledgements. We thank S. Goodbody, Z. Ghahramani and M. Pembrey for comments on the manuscript. This project was supported by the MRC, the Wellcome Trust, Help a Child to See charity, the IRIS fund, the Child Health Research Appeal Trust and the Royal Society.

Correspondence and requests for materials should be addressed to C.M.H.

Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons

J. M. Hupé, A. C. James*, B. R. Payne*, S. G. Lomber*, P. Girard & J. Bullier

Cerveau et Vision INSERM 371, 18 avenue du Doyen Lépine, 69675 Bron Cédex, France

A single visual stimulus activates neurons in many different cortical areas. A major challenge in cortical physiology is to understand how the neural activity in these numerous active zones leads to a unified percept of the visual scene. The anatomical basis for these interactions is the dense network of connections that link the visual areas. Within this network, feedforward connections transmit signals from lower-order areas such as V1 or V2 to higher-order areas. In addition, there is a dense web of feedback connections which, despite their anatomical prominence^{1–4}, remain functionally mysterious^{5–8}. Here we show, using reversible inactivation of a higher-order area (monkey area V5/MT), that feedback connections serve to amplify and focus activity of neurons in lower-order areas, and that they are important in the differentiation of figure from ground, particularly in the case of stimuli of low visibility. More specifically, we show that feedback connections facilitate responses to objects moving within the classical receptive field; enhance suppression evoked by background stimuli in the surrounding region; and have the strongest effects for stimuli of low salience.

We recorded single units and multiunits (114 single units and 54 multiunits) in areas V1, V2 and V3 of anaesthetized and paralysed macaque monkeys. To study the role of feedback connections from area V5, a small region of the superior temporal sulcus (STS) containing this area was reversibly inactivated by cooling; we then compared the neuronal responses before, during and after STS inactivation. We used visual stimuli consisting of an optimally orientated bar moved across the centre of the receptive field on a background of irregularly distributed, half light and half dark, but lower luminance, square checks (Fig. 1d). In a sequence of interleaved stimulus conditions, the bar and background moved one at a time, or together, in the preferred direction for the cell or its opposite.

Figure 1a–c illustrates a spectrum of effects of the V5 inactivation for single neurons recorded in areas V1, V2 and V3, and stimulated by a bright bar moving in front of a stationary background of lower luminance contrast. A substantial and highly significant decrease in the response to the moving bar is observed in each case during V5 inactivation. Figure 1e, f gives the population data. It is clear that diminution of responses is by far the most frequent effect of V5 inactivation, as observed before for other feedback connections^{6,8}. Of the total sample of sites tested, 33% showed a significant decrease ($P < 0.01$) and 6.5% an increase. Similar effects were observed in infragranular and supragranular layers. No effect was observed in layer 4C of area V1.

The role of feedback connections in figure–ground discrimination was suggested to us when we found that the strength of the effect of V5 inactivation depended on the visibility of the stimuli used for testing neurons in area V3, an area that receives a particularly large feedback input from MT/V5 (ref. 9). We com-

* Present addresses: Mammalian Neurobiology and Reproduction, RSBS ANU, GPO Box 475, Canberra, ACT 2601, Australia (A.C.J.); Laboratory of Visual Perception and Cognition, Center for Advanced Biomedical Research, Boston University School of Medicine, 700 Albany Street, Massachusetts 02118, USA (B.R.P. and S.G.L.).

pared the effects of V5 inactivation on the responses of V3 neurons for different luminances of the central moving bar and the stationary background. We quantified the visibility of the bar against the background by a salience index that corresponds to the ratio of the contrasts of the central bar and the background. Figure 2 shows that the suppression of the response during cooling for V3 neurons is stronger for low-salience stimuli than for middle- and high-salience stimuli. For the low-salience stimuli, the bar is barely visible when both bar and background are stationary or moving coherently. The movement of the bar on the stationary background makes it clearly visible. Feedback from V5 could therefore provide the information on motion contrast to V3 neurons. As shown in Fig. 2 (right box), there were also response reductions in areas V1 and V2 for high-salience stimuli. It is important to mention that, for the whole sample in areas V1, V2 and V3, response increases were observed only for high-salience stimuli (not shown).

Because of the high degree of convergence of feedback connections¹⁰ and the larger receptive fields in higher-order areas, we suspected that feedback connections are involved in the integration of information concerning different parts of the visual field. We therefore compared the responses of neurons to the central bar moved on a stationary or moving background. As expected from the inhibitory interactions between centre and surround in receptive fields of many visual cortical neurons^{11–13}, the response to the bar and background when moved together was usually much weaker than when the bar moved on the stationary background (Fig. 3a, hatched histograms). For a number of V3 neurons, inactivating V5 had a differential effect on the responses to these two stimuli. When the central bar moved alone, the inactivation of area V5 decreased the response, as already shown in Fig. 1. In contrast, it enhanced the response when the central bar and background moved coherently (Fig. 3a, open histograms). In most cases, the response to the background moving alone was null or very small (Fig. 3a).

The salience was also found to be important in determining the strength of effects elicited by inactivation of area V5. When V5 was inactivated, there was a marked increase in the response to the bar

and background moved together (BM stimulus), which was significant at the population level only in the case of low-salience stimuli (Fig. 3b). This suggests that the suppression of the response induced by the moving background is under the influence of area V5. For the low-salience stimuli (Fig. 3c), when area V5 was inactivated, there was a substantial decrease in background suppression for all the neurons except one, as shown by the positions of most points well below the line of no change in suppression. In most cases (5/7), this decrease in background suppression corresponded to a statistically significant increase in the response to the BM stimulus. An effect on a smaller proportion of the neurons was observed for middle-salience cases and no effect was seen for the high-salience cases (Fig. 3d). Thus, feedback afferents from V5 specifically increase the surround-induced suppression of the centre-mechanism response in V3 neurons in the case of low-salience stimuli. Note that the bar contrast is not the determinant variable: when classifying the neurons with respect to bar contrast, the effects are no longer clustered (not shown).

Feedback projections from area V5 also contribute importantly to the direction selectivity of lower-order neurons when salience is low. As for the surround suppression, changes in direction selectivity in V3 neurons were observed for low-salience stimuli (increases as well as decreases in direction selectivity were observed; not shown) and no effect was observed for middle and high salience. These results suggest that, at least for low-salience stimuli, the deficits in discrimination of stimulus direction after lesions in area V5 (refs 14, 15) are not necessarily due solely to the elimination of the numerous direction-selective units found in this area.

The results show that cooling the depth of the STS decreases the responses of many neurons in areas V1, V2 and V3. In addition to direct controls (see Methods), several indirect arguments contradict the possibility of a direct effect of the cold on the recorded regions or the radiations fibres. (1) There was no statistically significant correlation between the incidence or the strength of the response decrease and the depth of recording or the temperature of the probe. (2) Neurons showing response increases or no changes were mixed

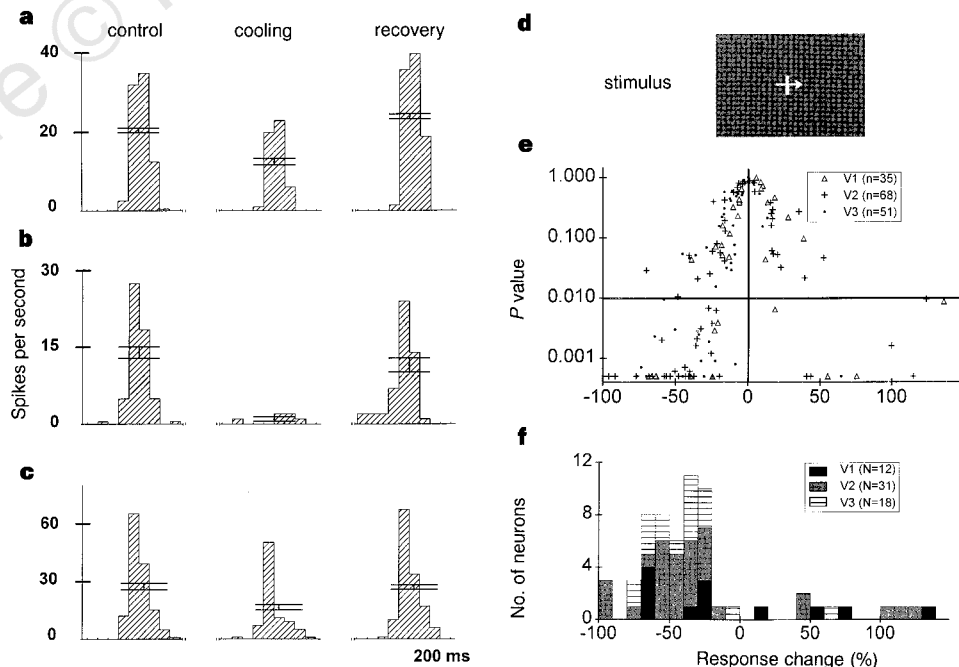


Figure 1 Effect of V5 inactivation on responses of neurons in areas V1, V2, V3. The stimulus was a light bar moving on a stationary background of lower contrast. **a–c**, Points and bars over each histogram represent the mean and s.e.m. Bin width = 50 ms. **a**, Area V1, 39% decrease of the response, case lca21. **b**, Area V2, 91% decrease, case kas11. **c**, Area V3, 40% decrease, case lck21. **d**,

Illustration of the stimulus. **e**, Scattergram of statistical significance (P -value) versus percentage change in response. The horizontal bar indicates the level of significance used ($P < 0.01$). **f**, Distribution histogram of the percentage change in response for significant effects ($P < 0.01$).

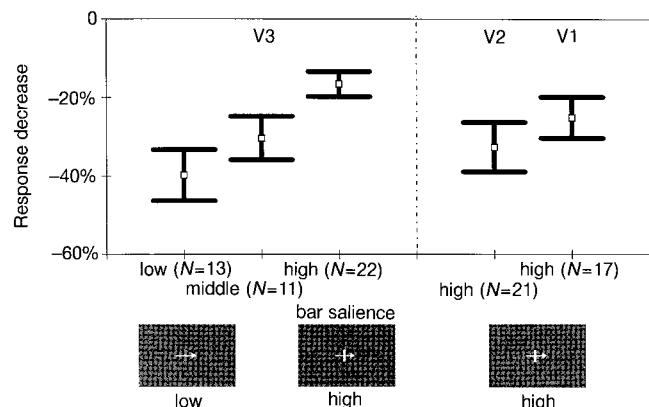


Figure 2 Effect of bar salience. Left box, effect on the response decreases due to V5 inactivation for V3 neurons stimulated with a moving bar on a stationary textured background (mean values $\pm$ s.e.m.; calculated for all the neurons showing response decreases). Strongest effects are observed for low salience of the bar (Kruskal-Wallis test, $P = 0.003$). A similar, but nonsignificant trend was observed in areas V1 and V2. Right box, values of response reductions for high-salience stimuli in areas V1 and V2. The reductions at high salience were not significantly different in the three areas ($P = 0.13$ in a Kruskal-Wallis test). The salience is defined as the ratio of the contrasts of the bar and the background. Low salience, 1–7; medium salience, 7–15; high salience, >15.

among neurons showing response decreases in the same penetrations. (3) Most effects were stimulus specific: besides the mentioned dependence on salience, we found that for most neurons showing a significant response decrease for the moving central bar, there was an increase or no change in response to at least one other stimulus presented in an interleaved fashion (Fig. 3a).

It is clear, however, that areas of the V5 complex that surround V5, such as areas MST, FST and V4t, are partly affected by the cooling. We believe that the effects we observed were mostly due to inactivation of area V5, as this area provides the strongest contingent of feedback connections to areas V1, V2 and V3 (refs 2–4, 9, 16). This conclusion is supported by the observation that no effect was observed in another animal in which the probes were placed rostral to V5 in the STS.

Our results show that feedback connections from area V5 have a facilitatory effect on the responses of neurons in areas V1, V2 and V3 to a bar moved on a stationary background. This agrees with known anatomy showing feedback projections terminating mainly on spiny, excitatory neurons, presumably of pyramidal type, and much less so on sparsely spinous inhibitory neurons¹⁷. The boosting effects of feedback connections can be quite strong, particularly for low-salience stimuli: some neurons in V1, V2 and V3 are completely silenced in the absence of a feedback input from V5 (Fig. 1b, f). This means that the activity of a neuron in a given cortical area is not simply shaped by its feedforward inputs and the local network of horizontal connections, but depends crucially on the activity of neurons located in higher-order areas and transferred through feedback connections.

Although inactivation of feedforward^{18,19} as well as feedback connections produce mainly a decrease of neuronal responses, there are clear differences between the roles of the two pathways. For example, when V1 is inactivated, all activity is abolished in area V2, despite the fact that most neurons remain active in V5 (refs 18–20). In other words, in the absence of a drive from V1 through feedforward connections, feedback connections from V5 are unable to drive neurons in V2. Similar conclusions on feedforward/feedback differences have been reached from experiments using cats, where deactivation of the middle suprasylvian cortex, a possible homologue of V5, has little effect on neural activity revealed in areas 17 and 18 by 2-deoxyglucose (2DG)⁷. It is therefore likely that, at least in anaesthetized preparations, feedback connections act more as a gain enhancer, or gate, of activity already present, rather than as

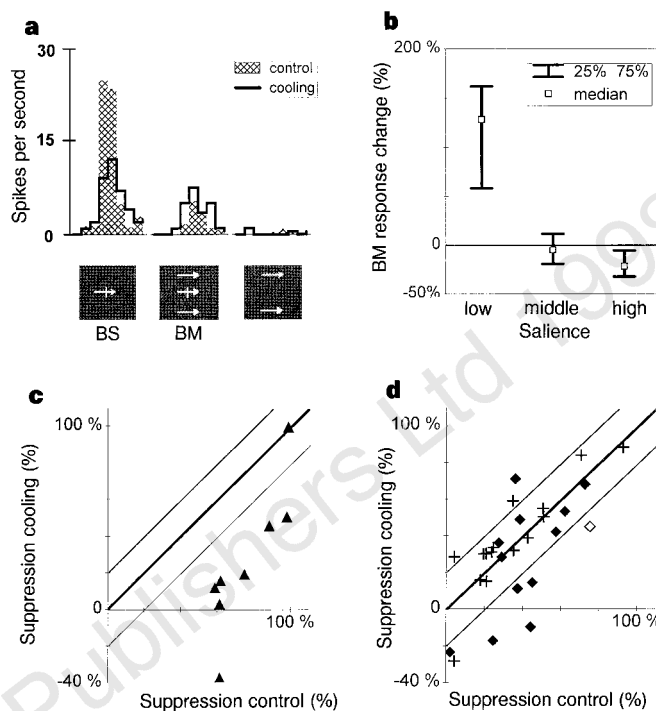


Figure 3 Effect of V5 inactivation on foreground-background interactions in V3 neurons. **a**, Responses before and during V5 cooling to a central bar moved on a stationary background (BS), to the same central bar moved coherently with the background (BM), and to the background alone. Case lcj21, area V3, salience 13.1, bin width 100 ms. **b**, Percentage change in response for the BM stimulus (bar and background moving together) during V5 inactivation. Only cases in which BS > BM activity were included. The difference between groups is highly significant ($P = 0.0086$ in a Kruskal-Wallis test; salience as defined in Fig. 2: low, $n = 8$; middle, $n = 13$; high, $n = 15$). **c**, Scattergram of the background-induced response suppression ($100 \times (BS - BM)/BS$) during V5 cooling versus control (mean value) for low salience. Central diagonal line indicates no change; the flanking lines indicate changes within 20% which include practically all the points in a similar scattergram between the two control runs. Values below the line of no change (central dark line) correspond to a decrease in response suppression. **d**, The same as (c) but for middle-salience stimuli (filled diamonds) and high-salience stimuli (+ symbols). Open symbol corresponds to example in (a). The cooling-induced changes in suppression for different groups of salience was highly significant (Kruskal-Wallis test, $P = 0.0004$; there was no significant effect between the controls: Kruskal-Wallis test, $P = 0.45$).

an activator of otherwise silent neurons, a role presumably reserved for feedforward connections.

Inactivation in V5 usually had opposing effects on figure-ground stimuli of low salience, decreasing the response to the central bar and increasing the response to the bar and background moved coherently. Thus, feedback connections may act in a push-pull fashion, amplifying the response to the optimal stimulus for the centre mechanism and decreasing that to stimuli activating centre and surround (Fig. 3a). These push-pull interactions are at their strongest for low-salience stimuli. This interesting specificity implies that feedback projections serve to improve the visibility of features that activate the receptive field centre in the stimulus and may thus contribute to figure-ground segregation, breaking of camouflage, and psychophysically demonstrated 'pop-out' effects.

Methods

Physiology. Recordings were obtained from two anaesthetized, paralysed cynomolgus monkeys. Procedures were similar to those reported earlier²¹. For analgesia, we replaced fentanyl by sufentanil (usually $4 \mu\text{g kg}^{-1} \text{h}^{-1}$). A spike discriminator (MSD from Alpha Omega) was used to extract single units from our recordings and to monitor neuronal identity during periods of control,

cooling and recovery. The receptive fields of recorded neurons were located in the central 4° of visual field. Electrolytic lesions were used to aid reconstruction of electrode tracts on histological sections stained for Nissl substance, cytochrome oxidase and Cat301. We relied principally on Cat301 labelling²² to identify V3 and MT/V5.

Cooling. The general procedure was similar to that described earlier²³. Several months before the experiment, two probes, consisting of a 7 mm × 3 mm loop of hypodermic tubing, were placed side-by-side in the depth of the STS at the level of area V5, ipsilateral to the recordings. The proper placement of the probes was verified post-mortem on histological sections stained as above. Cooled methanol was circulated simultaneously through both probes. Probe temperatures (mean value 7.2 °C, range 2.5–12 °C) were monitored by thermocouples at the base of the cooling loops.

Runs usually lasted for 3.5 min. Two control runs were done before cooling was applied. Recording during the cooling run was started once the probes reached constant temperature, a few minutes after the chilled methanol began circulating. After a pause of 15–30 min after the end of the cooling run, one or several recovery runs were done.

It is important to determine to what extent the tissue is affected by the cold. It was shown earlier that the average temperature gradient created by the probe is 10 deg mm⁻¹ and that neuropil beyond 3–3.5 mm from the cooling probe is at physiological temperature (Fig. 1 of Lomber *et al.*²³). This was confirmed by 2DG mapping showing that the metabolic activity of cortex located 3 mm away from the cooling probe was totally unaffected by the cooling⁷. From earlier anatomical publications², we measured the following shortest distances between retinotopically corresponding regions in STS and V1, V2, V3 and V4: 7, 5.7, 7 and 3.4 mm. Thus it is highly unlikely that these areas are affected by the cold. This was confirmed by measuring in two instances the temperature from the surface of area 17 down to a depth of 8 mm (corresponding to our recordings in V3). In one experiment there was no change in temperature throughout the cortical depth and in the other case the temperature dropped by less than 2 °C when the probes were cooled.

Furthermore, because fibres of the optic radiations travel 1–2 mm below the STS grey matter²⁴, even for the lowest temperature of our probes (2.5 °C), these fibres did not reach temperatures lower than 27 °C. Because fibres have a blocking temperature at least 10 deg lower than cell bodies that are inactivated below 20 °C^{25,26}, it seems very unlikely that direct blocking of LGN fibres could explain our results.

Visual stimuli. Visual stimuli were presented on a computer monitor driven by a Truevision Vista Board under the control of a Matlab program. Stimuli were presented twenty times in an interleaved fashion. The bar was approximately optimized in size and velocity for each neuron site studied. Orientation was optimized to within 15° by measurement of an orientation tuning curve. The background covered a field that was 12.7° wide and 8.4° high and that comprised randomly distributed checks of a size equal to the bar width. The background had the appearance of a set of bars of variable length and similar orientation as the central bar, which was only visible against the background if it differed in contrast or relative movement. One set of luminance values was used at each site, with mean background luminance, L_0 , being in the range 9–24 Cd m⁻². The contrast of the bar relative to this luminance, $C_{\text{bar}} = (I_{\text{bar}} - L_0)/L_0$, was in the range 0.72–12.6. The contrast of the light background checks relative to L_0 , $C_{\text{check}} = (L_{\text{check}} - L_0)/L_0$ (equal to the Michelson contrast of the light–dark checks) was in the range 0.06–0.98. The salience is equal to $C_{\text{bar}}/C_{\text{check}}$. Selection of different luminance combinations for the bar and background was not systematic.

Data processing. To limit the presence of false positive results due to poor stationarity of the cortex, data were analysed only when the responses to the first two control runs were not significantly different ($P > 0.01$). Responses were measured as the mean number of spikes over the stimulation period for 20 repetitions of the stimulus. Because of the non-gaussian distribution of the data and occasional changes in variance between conditions, we used a bias-corrected and accelerated-bootstrap Student-*t* procedure²⁷ to assess the statistical significance of differences in the mean number of spikes across the different runs.

Received 8 April; accepted 12 June 1998.

- Salin, P.-A. & Bullier, J. Corticocortical connections in the visual system: structure and function. *Physiol. Rev.* **75**, 107–154 (1995).
- Kennedy, H. & Bullier, J. A double-labelling investigation of the afferent connectivity to cortical areas

- V1 and V2 of the macaque monkey. *J. Neurosci.* **5**, 2815–2830 (1985).
- Perkel, D. J., Bullier, J. & Kennedy, H. Topography of the afferent connectivity of area 17 of the macaque monkey: a double-labelling study. *J. Comp. Neurol.* **253**, 374–402 (1986).
- Shipp, S. & Zeki, S. The organization of connections between areas V5 and V2 in macaque monkey visual cortex. *Eur. J. Neurosci.* **1**, 333–354 (1989).
- Alonso, J. M., Cudeiro, J., Pérez, R., Gonzales, F. & Acuna, C. Influence of layer 5 of area 18 of the cat visual cortex on responses of cells in layer 5 of area 17 to stimuli of high velocity. *Exp. Brain Res.* **93**, 363–366 (1993).
- Sandell, J. H. & Schiller, P. H. Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *J. Neurophysiol.* **48**, 38–48 (1982).
- Vanduffel, W., Payne, B. R., Lomber, S. G. & Orban, G. A. Functional impact of cerebral connections. *Proc. Natl Acad. Sci. USA* **94**, 7617–7620 (1997).
- Mignard, M. & Malpeli, J. G. Paths of information flow through visual cortex. *Science* **251**, 1249–1251 (1991).
- Felleman, D. J., Burkhalter, A. & Van Essen, D. C. Cortical connections of areas V3 and VP of macaque monkey extrastriate visual cortex. *J. Comp. Neurol.* **379**, 21–47 (1997).
- Salin, P. A., Girard, P., Kennedy, H. & Bullier, J. The visuotopic organization of corticocortical connections in the visual system of the cat. *J. Comp. Neurol.* **320**, 415–434 (1992).
- Nelson, J. I. & Frost, B. J. Orientation-selective inhibition from beyond the classic visual receptive field. *Brain Res.* **139**, 359–365 (1978).
- Li, C.-Y. & Li, W. Extensive integration of field beyond the classical receptive field of cat's striate cortical neurons—classification and tuning properties. *Vision Res.* **34**, 2337–55 (1994).
- Levitt, J. B. & Lund, J. S. Contrast dependence of contextual effects in primate visual cortex. *Nature* **387**, 73–76 (1997).
- Pasternak, T. & Merigan, W. H. Motion perception following lesions of the superior temporal sulcus in the monkey. *Cerebr. Cortex* **4**, 247–259 (1994).
- Newsome, W. T. & Pare, E. B. A selective impairment of motion processing following lesions of the middle temporal visual area (MT). *J. Neurosci.* **8**, 220–2211 (1988).
- Shipp, S. & Zeki, S. The organization of connections between areas V5 and V1 in macaque monkey visual cortex. *Eur. J. Neurosci.* **1**, 308–331 (1989).
- Johnson, R. R. & Burkhalter, A. Microcircuitry of forward and feedback connections within rat visual cortex. *J. Comp. Neurol.* **368**, 383–398 (1996).
- Girard, P., Salin, P. A. & Bullier, J. Response selectivity of neurons in area MT of the macaque monkey during reversible inactivation of area V1. *J. Neurophysiol.* **67**, 1–10 (1992).
- Bullier, J., Girard, P. & Salin, P. A. in *Primary Visual Cortex in Primates* (eds Peters, A. & Rockland, K. S.) 301–330 (Plenum Pub. Corp., 1994).
- Rodman, H. R., Gross, C. G. & Albright, T. D. Afferent basis of visual response properties in area MT of the macaque: I. Effects of striate cortex removal. *J. Neurosci.* **9**, 2033–2050 (1989).
- Nowak, L. G., Munk, M. H. J., Girard, P. & Bullier, J. Visual Latencies in Areas V1 and V2 of the Macaque Monkey. *Vis. Neurosci.* **12**, 371–384 (1995).
- DeYoe, E. A., Hockfield, S., Garren, H. & Essen, D. C. V. Antibody labeling of functional subdivisions in visual cortex: Cat-301 immunoreactivity in striate and extrastriate cortex of the macaque monkey. *Vis. Neurosci.* **5**, 67–81 (1990).
- Lomber, S. G., Payne, B. R. & Cornwell, P. Learnign and recall of form-discriminations during reversible cooling deactivation of ventral-posterior suprasylvian cortex in the cat. *Proc. Natl Acad. Sci. USA* **93**, 1654–1658 (1996).
- Polyak, S. *The Vertebrate Visual System* (Chicago Univ. Press, 1957).
- Bénita, M. & Condé, H. Effects of local cooling upon conduction and synaptic transmission. *Brain Res.* **36**, 133–151 (1972).
- Brooks, V. B. Study of brain function by local, reversible cooling. *Rev. Physiol. Biochem. Pharmacol.* **95**, 1–109 (1983).
- Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).

Acknowledgements. We thank Henry Kennedy, Ken Knoblauch, Jonathan Levitt and Matthias Munk for fruitful discussions and a careful reading of the manuscript, and Susan Hockfield for providing us the CAT301 antibody. This work was supported by NIH, NATO, Biomed, and GIS contract 'Perception visuelle'.

Correspondence and requests for material should be addressed to J.B. (e-mail: bullier@lyon151.inserm.fr).

Stress and glucocorticoids impair retrieval of long-term spatial memory

Dominique J.-F. de Quervain*, Benno Roozendaal* & James L. McGaugh*†

* Center for the Neurobiology of Learning and Memory, and † Department of Psychobiology, University of California, Irvine, California 92697-3800, USA

Extensive evidence from animal and human studies indicates that stress and glucocorticoids influence cognitive function^{1–11}. Previous studies have focused exclusively on glucocorticoid effects on acquisition and long-term storage of newly acquired information. Here we report that stress and glucocorticoids also affect memory retrieval. We show that rats have impaired performance in a water-maze spatial task after being given footshock 30 min before retention testing but are not impaired when footshock is given 2 min or 4 h before testing. These time-dependent effects on retention performance correspond to the circulating corticoster-

one levels at the time of testing, which suggests that the retention impairment is directly related to increased adrenocortical function. In support of this idea, we find that suppression of corticosterone synthesis with metyrapone blocks the stress-induced retention impairment. In addition, systemic corticosterone administered to non-stressed rats 30 min before retention testing induces dose-dependent retention impairment. The impairing effects of stress and glucocorticoids on retention are not due to disruption of spatial navigation *per se*. Our results indicate that besides the well described effects of stress and glucocorticoids on acquisition and consolidation processes, glucocorticoids also affect memory retrieval mechanisms.

Young adult male Sprague–Dawley rats (aged 10–12 weeks) from Charles River Laboratories were trained in a single session of eight trials in a water maze to find a submerged platform located in a fixed position¹². Memory for the location of the platform was tested 24 h later using a 60-s free-swim probe trial. In the first experiment, rats received three footshocks (0.8 mA for 1 s with a 5-s intershock

interval) either 2 min, 30 min or 4 h before the retention probe trial. Non-stressed controls spent significantly more time in the target quadrant than in the opposite quadrant ($P < 0.01$), indicating memory for the location of the platform (Fig. 1a). Retention was impaired when the animals were tested 30 min ($P < 0.01$), but not 2 min or 4 h, after the footshock stimulation. Figure 1b shows representative examples of the swimming pattern of a non-stressed control and that of a rat given footshock stimulation 30 min before testing. Control rats displayed focused searching for the platform in the correct location, whereas stressed rats (30-min group) showed random swimming patterns. The four groups did not differ in total pathlengths during the probe trial ($F_{3,36} = 0.04$, $P > 0.98$). These results indicate that the stress impaired specific aspects of spatial performance and that the retention performance on the probe trial was not affected by any gross disturbances in motor performance.

These results suggest that stress induces time-dependent impairment of memory retrieval. Stress initiates a cascade of physiological events, including high circulating levels of catecholamines and

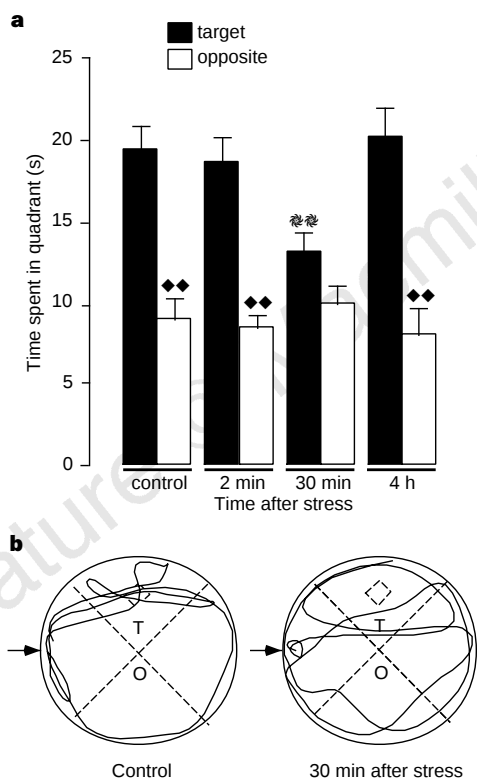


Figure 1 Effects of pre-testing exposure to footshock on retention performance. **a**, Time (mean $\pm$ s.e.m.) in seconds spent in target quadrant (T; filled bars) and opposite quadrant (O; open bars) for the four experimental groups. A two-way ANOVA showed significant treatment ($F_{3,36} = 3.68$, $P < 0.05$) and quadrant effects ($F_{1,36} = 55.86$, $P < 0.0001$). **, $P < 0.01$ compared with non-stressed controls; ♦♦, $P < 0.01$ compared with time in target quadrant. Sample size is ten individuals per group. **b**, Representative swimming paths of a non-stressed control and an animal given footshock stress 30 min before testing. Arrow indicates starting placement.

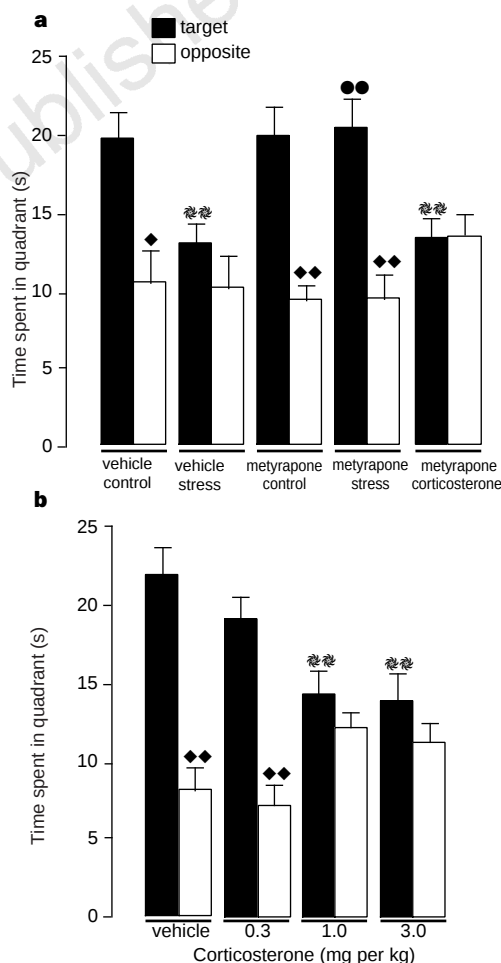


Figure 2 Effects of pre-testing manipulation of glucocorticoid levels on retention performance. Time (mean $\pm$ s.e.m.) in seconds spent in target quadrant (filled bars) and opposite quadrant (open bars). **a**, Effect of metyrapone on stress-induced impairment in retention performance. A two-way ANOVA showed significant treatment ($F_{4,42} = 4.94$, $P < 0.005$) and quadrant effects ($F_{1,42} = 22.01$, $P < 0.0001$). **, $P < 0.01$ compared with vehicle-control group; ♦♦, $P < 0.05$; ♦♦♦, $P < 0.01$ compared with time in target quadrant. **b**, Effects of pre-testing corticosterone administration on retention performance. A two-way ANOVA showed significant quadrant effects ($F_{1,36} = 35.47$, $P < 0.0001$) and treatment $\times$ quadrant interaction effects ($F_{3,36} = 5.63$, $P < 0.01$). **, $P < 0.01$ compared with vehicle group; ♦♦♦, $P < 0.01$ compared with time in target quadrant. Sample size is between nine and ten individuals per group.

glucocorticoids. Catecholamines are released rapidly and typically return to baseline levels within 10 min, but the glucocorticoid response is much slower¹³. The finding that plasma levels of corticosterone, from blood collected immediately after the probe trial, were elevated in the impaired 30-min group ($P < 0.01$), but not in any of the other stress groups (Table 1a), suggests that increased adrenocortical function induced by the stressor may have disrupted memory retrieval. To examine this possibility, rats were given metyrapone, a drug that reduces the synthesis of corticosterone by inhibiting the 11β -hydroxylation reaction in the adrenal glands. Metyrapone (50 mg per kg, subcutaneous (s.c.) injection; Sigma) was dissolved in a vehicle containing 40% polyethylene glycol and 60% saline. Consistent with previous findings¹⁴, metyrapone injected 40 min before footshock had no effect on baseline corticosterone levels, but blocked the stress-induced increase in plasma corticosterone ($P < 0.01$; Table 1b). Metyrapone also blocked the stress-induced retention impairment ($P < 0.01$, Fig. 2a). Other possible actions of metyrapone were excluded as metyrapone did not block the retention impairment in rats given corticosterone (3.0 mg per kg, s.c.) 30 min before testing. Neither of the treatments altered swimming pathlengths ($F_{4,42} = 0.61$, $P > 0.64$).

These results suggest that activation of the hypothalamic–pituitary–adrenal cortex (HPA) axis mediates the impairing effects of footshock stress on retention for spatial memory and that a selective blockade of the HPA axis is sufficient to block these impairing effects. To investigate further the putative role of glucocorticoids on memory retrieval, non-stressed rats were given corticosterone (0.3, 1.0 or 3.0 mg per kg, s.c.; Sigma) 30 min before the probe trial (Fig. 2b). Corticosterone was dissolved in a vehicle containing 5% ethanol in saline. As expected, the systemic corticosterone injections induced dose-dependent increases in plasma corticosterone levels (Table 1c). The two higher doses of corticosterone induced retention impairment comparable to that seen after footshock ($P < 0.01$ for both doses). Again, the groups did not differ in total pathlengths ($F_{3,36} = 1.70$, $P > 0.18$). Results of a previous study indicated that intracerebroventricular administration of the specific antagonists for adrenal steroid receptors to non-stressed rats shortly before retention testing did not affect retention performance in a spatial water-maze task¹⁵. Together, these findings suggest that stress levels of glucocorticoids may impair memory retrieval, whereas basal levels do not exert tonic inhibitory effects on memory retrieval. Moreover, because the water-maze training procedure is very stressful and induces high levels of corticosterone, the present effects cannot be explained by state dependency.

Many nonspecific factors affecting sensory, attentional and motor

processes influence spatial navigation. As indicated above, motor performance was not affected by any of the experimental manipulations. To examine further whether the effects were due selectively to influences on specific aspects of memory retrieval, footshock, corticosterone (3.0 mg per kg, s.c.) or a vehicle injection were given to naive animals 30 min before a water-maze training session. A two-way analysis of variance (ANOVA), with trials as repeated measures, showed that all three groups learned the task and that they did not differ in escape latencies ($F_{2,189} = 0.08$, $P > 0.92$). In addition, the group did not differ in performance on a probe trial given 20 s after the last training trial ($F_{2,27} = 0.08$, $P > 0.93$). These findings indicate that the impairing effects of stress and glucocorticoids on retention performance are not due to nonspecific effects on spatial navigation. These data further indicated that immediate recall is not affected by the treatments.

Evidence summarized by Lupien and McEwen indicates that glucocorticoids have multiple and often conflicting effects on memory function⁶. These authors stressed that in order to interpret correctly the effects of glucocorticoids on cognition, it is critical to dissociate the effects on the different memory phases. The time-limited effects of stress and glucocorticoids on memory consolidation follow an inverted-U-shape dose–response relationship: extreme low and high levels may impair consolidation^{1–4,6–8}, but intermediate doses enhance memory^{5,6,9–11}, and activation of glucocorticoid receptors seems to be a prerequisite for the long-term storage of information. To our knowledge, our report is the first to demonstrate that once memories are consolidated, the efficacy or accuracy of the information retrieved remains vulnerable to glucocorticoids at the time of recall. As the retrieval impairment was found as soon as 30 min after drug injections, which is too fast to consider the classic genomic action via intracellular receptors¹⁶, it is likely that the effects may involve an interaction with neurotransmitter mechanisms^{17–19}. Alternatively, other rapid, non-genomic steroid actions have been described via surface receptors linked to GABA receptors²⁰. Whatever the exact mechanism, glucocorticoid influences on retrieval mechanisms seem to be exclusively adverse. In fact, the same treatments that enhance memory consolidation when administered immediately post training^{9–11}, impaired memory retrieval in the present study when given shortly before testing. Although our experiments examined the role of stress and glucocorticoids on retrieval in only a relatively stressful task, the findings reinforce the conclusion that the nature of the effects of glucocorticoids on cognitive performance depends on the specific mnemonic processes examined. Further research is needed to determine whether comparable effects are found with less stressful tasks and with other species. Our results are important in relation to observations in human subjects that indicate that diurnal fluctuations in cognitive performance are negatively correlated with diurnal variations in plasma cortisol levels²¹. Our findings may be of even greater importance in considering experiments that examine the effects, on memory function, of chronic stress or sustained elevated levels of glucocorticoids, as well as in clinical conditions of human patients with hypercortisolism. In contrast to acute treatments, chronic exposure to stress or glucocorticoids seems to induce only cognitive deficits. Traditionally, the impairments seen under those conditions have been interpreted (either implicitly or explicitly) as being due solely to a disruption of memory consolidation. On the basis of the present findings, however, it seems likely that at least some of the effects may be due to impaired memory retrieval. □

Methods

The water maze was a circular galvanized steel tank, 1.83 m in diameter, 0.58 m in height, filled with water (27 °C) to a depth of 20 cm. For the training, a submerged Perspex platform (20 cm × 25 cm) was placed 25 cm away from the edge in a fixed location. The single training session consisted of eight trials with four different starting positions that were equally distributed around the perimeter of the maze. After mounting the platform, the animals were allowed

Table 1 Effects on plasma corticosterone levels

Treatment	Plasma corticosterone ($\mu\text{g dl}^{-1}$)
(a) Footshock stress	
Control	2.3 ± 0.2
2 min	6.1 ± 1.7
30 min	11.2 ± 4.0*
4 h	4.3 ± 0.8
(b) Metyrapone	
Vehicle control	4.1 ± 1.3
Vehicle stress	13.8 ± 4.9†
Metyrapone control	4.8 ± 1.1
Metyrapone stress	4.7 ± 0.5‡
Metyrapone corticosterone	29.1 ± 5.8†
(c) Corticosterone	
Vehicle	3.3 ± 0.7
0.3 mg kg ⁻¹	6.7 ± 1.5
1.0 mg kg ⁻¹	12.9 ± 2.6§
3.0 mg kg ⁻¹	31.0 ± 6.5

a, Plasma corticosterone levels (mean ± s.e.m.) of groups in Fig. 1a. A one-way ANOVA revealed a significant treatment effect ($F_{3,36} = 3.00$, $P < 0.05$). * $P < 0.01$ compared with non-stressed controls. **b**, Plasma corticosterone levels of groups in Fig. 2a. A one-way ANOVA revealed a significant group effect ($F_{4,36} = 9.52$, $P < 0.0001$). † $P < 0.01$ compared with vehicle-control group; ‡ $P < 0.01$ compared with vehicle-stress groups. **c**, Plasma corticosterone levels of groups in Fig. 2b. A one-way ANOVA showed a significant group effect ($F_{3,36} = 13.00$, $P < 0.0001$). § $P < 0.05$; || $P < 0.01$ compared with vehicle group. Sample size is between 8 and 10 individuals per group.

to remain there for 20 s, and were then placed in a holding cage for 30 s until the start of the next trial. After completion of training, the animals returned to their home cages until retention testing 24 h later. On the day of testing, the animals were retained in a waiting room for 1 h before exposure to footshock or systemic injection of corticosterone. Footshock stress was administered in an inhibitory avoidance apparatus (as described elsewhere⁹) in an adjacent, sound-attenuated room. The rats were placed in the starting compartment and, after they stepped completely into the shock compartment, the door between the compartments was closed and a series of three footshocks (0.8 mA for 1 s with a 5-s intershock interval) was administered. After the last footshock, the rats were retained in the shock compartment for 15 s and then returned to their home cages. The animals were assigned randomly to one of the experimental groups. The probe trial consisted of a 60-s free-swim period without a platform and was recorded on video tape for later analysis. The rat was placed in the tank at the arrow (Fig. 1b), a position that was equal in distance to the imaginary target quadrant (T) and opposite quadrant (O). Training and testing was conducted between 10:00 and 15:00 h.

For plasma corticosterone determination, the animals were decapitated immediately after the probe trial and trunk blood was collected in heparinized (500 IU ml⁻¹) tubes and stored on ice. After centrifugation at 5000 r.p.m. for 10 min, the supernatant was stored at -50 °C until assay. Corticosterone plasma concentrations were determined by radioimmunoassay using a highly specific antibody (B3-163, Endocrine Sciences, Tarzana, California) and ³H-corticosterone tracer. Coefficients of variation within and between assays were less than 10%.

Retention performance data were analysed with a two-way ANOVA with treatment as between-subject variable and quadrant as within-subject variable. The ANOVAs were followed by either Fisher's tests for between-subject comparisons or paired *t*-tests for within-subject comparisons. Plasma corticosterone levels were analysed with a one-way ANOVA followed by Fisher's tests. A probability of less than 0.05 was considered significant.

Received 24 March; accepted 18 June 1998.

- Conrad, C. D., Galea, L. A. M., Kuroda, Y. & McEwen, B. S. Chronic stress impairs spatial memory on the Y maze, and this effect is blocked by tianeptine pretreatment. *Behav. Neurosci.* **110**, 1321–1334 (1996).
- Kirschbaum, C., Wolf, O. T., May, M., Wiplich, W. & Hellhammer, D. H. Stress- and treatment-induced elevations of cortisol levels associated with impaired declarative memory in healthy adults. *Life Sci.* **58**, 1475–1483 (1996).
- Luine, V. N., Spencer, R. L. & McEwen, B. S. Effects of chronic corticosterone ingestion on spatial memory performance and hippocampal serotonergic function. *Brain Res.* **616**, 65–70 (1993).
- Luine, V., Villegas, M., Martinez, C. & McEwen, B. S. Repeated stress causes reversible impairments of spatial memory performance. *Brain Res.* **639**, 167–170 (1994).
- Luine, V., Martinez, C., Villegas, M., Magarinos, A. M. & McEwen, B. S. Restraint stress reversibly enhances spatial memory performance. *Physiol. Behav.* **59**, 27–32 (1996).
- Lupien, S. J. & McEwen, B. S. The acute effects of corticosteroids on cognition: integration of animal and human model studies. *Brain Res. Rev.* **24**, 1–27 (1997).
- Lupien, S. J. *et al.* Stress-induced declarative memory impairment in healthy elderly subjects: relationship to cortisol reactivity. *J. Clin. Endocrinol. Metabol.* **82**, 2070–2075 (1997).
- Newcomer, J. W., Craft, S., Hershey, T., Askins, K. & Bardgett, M. E. Glucocorticoid-induced impairment in declarative memory performance in adult humans. *J. Neurosci.* **14**, 2047–2053 (1994).
- Rooszendaal, B. & McGaugh, J. L. Amygdaloid nuclei lesions differentially affect glucocorticoid-induced memory enhancement in an inhibitory avoidance task. *Neurobiol. Learn. Memory* **65**, 1–8 (1996).
- Sandi, C., Loscertales, M. & Guaza, C. Experience-dependent facilitating effect of corticosterone on spatial memory formation in the water maze. *Eur. J. Neurosci.* **9**, 637–642 (1997).
- Shors, T. J., Weiss, C. & Thompson, R. F. Stress-induced facilitation of classical conditioning. *Science* **257**, 537–539 (1992).
- Morris, R. Developments of a water-maze procedure for studying spatial learning in the rat. *J. Neurosci. Meth.* **11**, 47–60 (1984).
- De Boer, S. F., Koopmans, S. J., Slangen, J. L. & Van der Gugten, J. Plasma catecholamine, corticosterone and glucose responses to repeated stress in rats: Effects of interstressor interval length. *Physiol. Behav.* **47**, 1117–1124 (1990).
- Rooszendaal, B., Bohus, B. & McGaugh, J. L. Dose-dependent suppression of adrenocortical activity with metyrapone: Effects on emotion and memory. *Psychoneuroendocrinol.* **21**, 681–693 (1996).
- Oitzl, M. S. & De Kloet, E. R. Selective corticosteroid antagonists modulate specific aspects of spatial orientation learning. *Behav. Neurosci.* **106**, 62–71 (1992).
- McEwen, B. S. Glucocorticoid-biogenic amine interactions in relation to mood and behavior. *Horm. Behav.* **28**, 396–405 (1994).
- De Kloet, E. R. Brain corticosteroid receptor balance and homeostatic control. *Front. Neuroendocrinol.* **12**, 955–164 (1991).
- Joels, M., Karst, H., Hesen, W. & Wadman, W. J. in *Brain Corticosteroid Receptors* (eds De Kloet, E. R., Azmitia, E. C. & Landfield, P. W.) 166–177 (Ann. NY Acad. Sci. 746, 1994).
- Thomas, D. N., Post, R. M. & Pert, A. in *Brain Corticosteroid Receptors* (eds De Kloet, E. R., Azmitia, E. C. & Landfield, P. W.) 467–469 (Ann. NY Acad. Sci. 746, 1994).
- Orchinik, M., Moore, F. L. & Rose, J. D. in *Brain Corticosteroid Receptors* (eds De Kloet, E. R., Azmitia, E. C. & Landfield, P. W.) 101–114 (Ann. NY Acad. Sci. 746, 1994).
- Monk, T. H. *et al.* Circadian rhythms in human performance and mood under constant conditions. *J. Sleep Res.* **6**, 9–18 (1997).

Acknowledgements. We thank R. M. Sapolsky and S. Brooke for performing the corticosterone assay, and B. Bohus for his comments on an early draft of this paper. The research was supported by a grant from NIMH.

Correspondence and requests for materials should be addressed to B.R. (e-mail: roozendaal@darwin.bio.ucl.edu).

The stomach is a source of leptin

André Bado*, Sandrine Levasseur†‡, Samir Attoub†‡, Stéphanie Kermorgant*, Jean-Pierre Laigneau*, Marie-Noëlle Bortoluzzi‡, Laurent Moizo*, Thérèse Lehy*, Michèle Guerre-Millo§, Yannick Le Marchand-Brustel‡ & Miguel J. M. Lewin*

* Institut National de la Santé et de la Recherche Médicale (INSERM) Unité 10, IFR2 Cellules Epithéliales, Hôpital Bichat, 170 boulevard Ney, 75018 Paris, France

‡ INSERM U145, Faculté de Médecine, avenue de Valombrose, 06107 Nice, France

§ INSERM U465, Institut des Cordeliers, 15, rue de l'Ecole de Médecine, 75006 Paris, France

† These authors contributed equally to this work

The circulating peptide leptin, which is the product of the *ob* gene¹, provides feedback information on the size of fat stores to central Ob receptors^{2,3} that control food intake and body-weight homeostasis^{4–6}. Leptin has so far been reported to be secreted only by adipocytes¹ and the placenta⁷. Here we show that leptin messenger RNA and leptin protein are present in rat gastric epithelium, and that cells in the glands of the gastric fundic mucosa are immunoreactive for leptin. The physiological function of this previously unsuspected source of leptin is unknown. However, both feeding and administration of CCK-8 (the biologically active carboxy-terminal end of cholecystokinin) result in a rapid and large decrease in both leptin cell immunoreactivity and the leptin content of the fundic epithelium, with a concomitant increase in the concentration of leptin in the plasma. These results indicate that gastric leptin may be involved in early CCK-mediated effects activated by food intake, possibly including satiety.

Oligonucleotides deduced from the cloned mouse *ob* gene¹ were used for screening of total RNA extracted from fundic

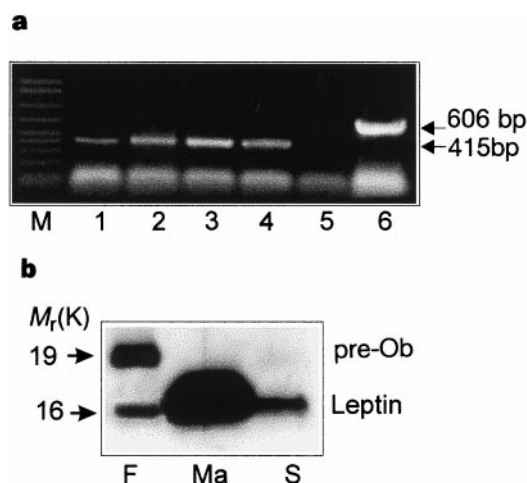


Figure 1 Leptin expression in Wistar rats. **a**, RT-PCR analysis of leptin mRNA: fundic epithelium (lane 1), mesenteric (lane 2), epididymal (lane 3) and perirenal (lane 4) adipose tissues; liver (lane 5); β -actin (lane 6); markers (lane M). Arrows indicate the expected size of the PCR products: 415 bp for *ob* and 606 bp for β -actin. **b**, Western blot of leptin protein: homogenate from fundic epithelium scrapings of Wistar rats (50 μ g proteins, lane F), mesenteric adipose tissue (20 μ g proteins, lane Ma) and serum (5 μ g proteins, lane S). Two leptin immunoreactive bands are seen in the fundic epithelium sample: 16K (leptin) and 19K (leptin precursor, pre-Ob). Note that the 19K band is absent from the adipose tissue and serum.

epithelium scrapings, adipose tissues of various origins (perirenal, epididymal and mesenteric) and liver. Screening was done using the polymerase chain reaction (PCR) after reverse transcription of RNA (RT-PCR). One *ob* PCR product of the expected size of 415 base pairs (bp) was found in fundic epithelium and all adipose tissue tested, but not in liver extract (Fig. 1a). The nucleotide sequence of the amplified product was determined, and was found to be identical to the mRNA product of the *ob* gene¹. This unexpected finding disagrees with a previous study that failed to detect leptin mRNA in the stomach¹. However, in this earlier work, total RNA from whole stomach was analysed by northern blotting and RT-PCR, whereas we used RNA that was selectively extracted from the fundic epithelium. Furthermore, we used a much larger amount of total RNA (4 µg compared with 0.1 µg) and more DNA denaturation–renaturation cycles (40 compared with 30 cycles). We did not detect any leptin mRNA in other sites in the gastrointestinal tract, including the small intestine, colon, rectum and pancreas (not shown). The presence of leptin protein in the fundic epithelium was investigated by immunoblot analysis (Fig. 1b). Two immunoreactive bands were observed: a 16K (relative molecular mass, 66,000) band corresponding to leptin and a 19K band, which was probably the leptin precursor¹. In agreement with previous reports, the 19K protein was absent from extracts of fat tissue. This variance between leptin-containing gastric cells and adipocytes presumably reflects a difference in protein processing and excretion. In addition, the amounts of gastric leptin in lean and obese Zucker rats were similar ($3.8 \pm 0.81 \text{ ng g}^{-1}$ compared with $4.1 \pm 0.83 \text{ ng g}^{-1}$ mucosa, respectively), although the latter had more than twice as much epididymal leptin per fat pad ($6.3 \pm 0.35 \text{ ng}$ compared with $22.3 \pm 1.5 \text{ ng}$, respectively; $P < 0.001$). This suggests an adipocyte specificity of leptin upregulation in these obese animals, as previously observed for other proteins^{8,9}.

We found that cells in the gastric epithelium possessed leptin immunoreactivity. They were localized in the lower half of the fundic glands, a site similar to that of the pepsinogen-secreting chief cells (Fig. 2). Leptin immunoreactivity in gastric cells decreased substantially after intraperitoneal administration of CCK-8. In parallel, the leptin content of the fundic epithelium decreased (Fig. 3). The amount of gastric leptin, measured by radioimmune assay (RIA), was slightly decreased by starvation but was not significantly different in rats that had been fasted for 18 h and control animals ($4.3 \pm 0.6 \text{ ng g}^{-1}$ compared with $5.4 \pm 0.4 \text{ ng g}^{-1}$, respectively). In contrast, the concentration of leptin in plasma declined sharply during an 18-h fast ($0.52 \pm 0.07 \text{ ng ml}^{-1}$ compared with $3.12 \pm 0.4 \text{ ng ml}^{-1}$ in rats fed *ad libitum*; $P < 0.001$). In animals that had fasted for 18 h, CCK-8 produced a dose-dependent decrease in the leptin content of the fundic epithelium. This decrease was significant at a dose of $10 \mu\text{g kg}^{-1}$ CCK-8 and maximal

within 15 min for $30 \mu\text{g kg}^{-1}$ CCK-8, whereas the plasma leptin level increased rapidly (Fig. 4).

The gastric hormone gastrin, whose structure is closely related to CCK, produced similar effects (not shown). Refeeding of fasted animals also resulted in a rapid and substantial decrease in gastric leptin content (Fig. 4). This decrease reached 66% after 15 min. Thereafter, there was a recovery phase, which probably reflected leptin neosynthesis in the gastric cells. Concomitantly, there was a threefold increase in the concentration of plasma leptin 15 min after the start of refeeding and a fourfold increase after 2 h to a value of 70% of the fed level (Fig. 4). This increase in plasma after two hours of refeeding is consistent with the reported stimulatory effect of refeeding on plasma leptin^{10–12} and adipocyte leptin mRNA^{13,14}. However, to our knowledge, this is the first report of leptin concentrations being rapidly modulated by both feeding and a peptide.

This finding cannot be explained by adipocyte secretion because adipose cells are not believed to store leptin. Furthermore, in agreement with another report¹⁵, CCK-8 did not stimulate leptin secretion from isolated rat adipocytes, nor did it affect leptin secretion from rat adipocytes transfected with a pCisOb complementary DNA (which directs overproduction of leptin) in comparison to the β_3 -receptor agonist BRL37344 which, as previously reported^{16,17}, inhibited leptin secretion (Table 1). In addition, 15 min after CCK injection, the leptin content of epididymal fat pads showed no significant change ($5.42 \pm 1.1 \text{ ng}$ compared with $6.70 \pm 0.8 \text{ ng}$). However, experiments with isolated, vascularized

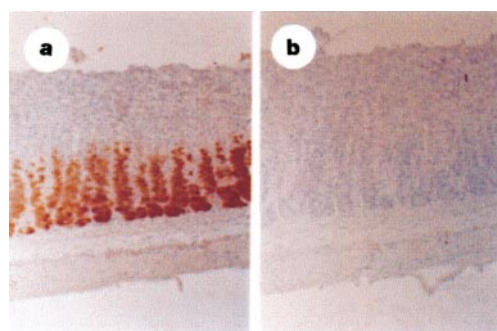


Figure 2 Immunostaining of leptin in rat fundic mucosa. **a**, Leptin immunoreactive cells are seen in the lower half of the fundic epithelial glands. **b**, Serial adjacent section, same region. No immunoreactivity is seen after adsorption of the antiserum with $10 \mu\text{g ml}^{-1}$ of leptin fragment [137–156].

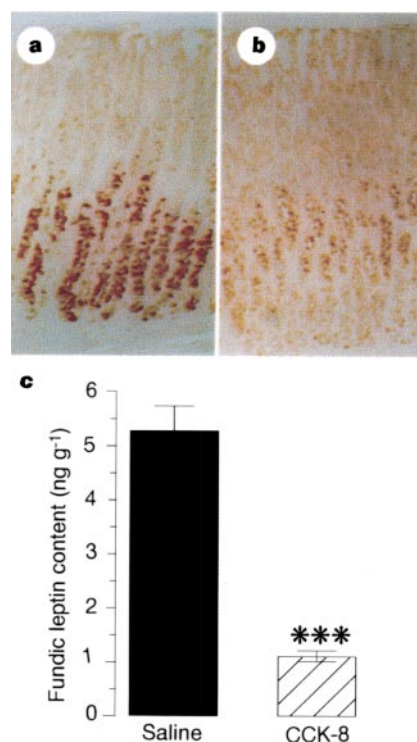


Figure 3 Effects of CCK-8 on immunostaining of leptin in rat fundic mucosa. Fed rats were given an intraperitoneal injection of saline (control) or CCK-8 ($300 \mu\text{g kg}^{-1}$). In these animals, gastric leptin immunostaining and leptin content were examined 15 min after injection. **a**, A saline-treated rat with cells possessing leptin immunoreactivity seen within the glands, in the lower half of mucosa. **b**, A decrease in immunostaining is evident 15 min after CCK-8 administration. **c**, Fundic content of leptin levels in saline- and the CCK8-treated rats 15 min after injection. Leptin was extracted from the fundic epithelium scrapings and assayed by RIA as described in Methods. Each column is the mean $\pm$ 1 s.e.m. for three rats in the control group and five rats in the CCK-treated group. Data were analysed by a Student *t*-test. ***, $P < 0.001$ compared with saline.

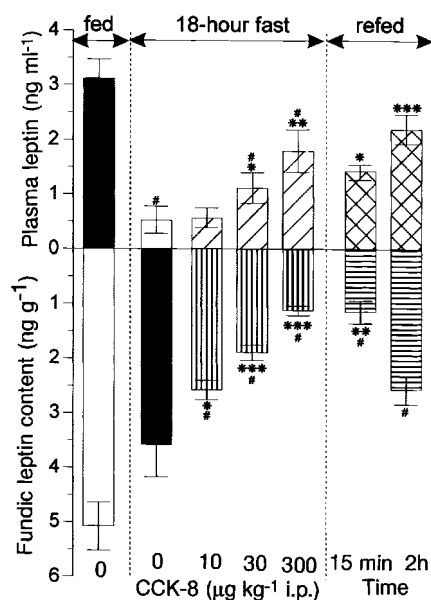


Figure 4 Effects of CCK-8 or feeding on plasma leptin and on fundic content. Wistar rats that were deprived of food for 18 h were injected intraperitoneally with saline and allowed to re-feed or with CCK-8 (Sigma). Blood was collected from the abdominal aorta at 15-min post-injection after anaesthesia. The stomach was removed and leptin extracted from fundic epithelium scrapings as described in Methods. Each column is the mean $\pm$ 1 s.e.m. for six to eight rats. Data were analysed by a Tukey-Kramer multiple comparisons test after a significant ANOVA. #, $P < 0.01$ compared with fed; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ compared with control (0).

stomach (Fig. 5) showed that CCK-induced depletion of gastric leptin content does result in leptin secretion into the blood compartment.

These observations indicate that the rapid increase in the concentration of plasma leptin seen in response to CCK involves the mobilization of a gastric leptin store. RIA determinations with epithelial scrapings ($5.4 \text{ ng leptin g}^{-1}$) indicate that this store was about 1.4 ng . However, this value is likely to be an underestimate because of incomplete scraping of the epithelium and proteolytic degradation during scraping. The increase in the plasma concentration was relatively small ($0.64 \pm 0.09 \text{ ng leptin ml}^{-1}$ for $30 \mu\text{g kg}^{-1}$ CCK-8), and therefore unlikely to have any large, direct effect on satiety control. However, the change could be sufficient to modify the effects of the synergistic interaction with CCK^{18,19}. Because of the very rapid degradation of CCK-8 (and CCK) in tissues and blood^{20,21}, there is a considerable difference between the injected doses and the resulting plasma concentration. Furthermore, CCK released by the intestinal I cells^{22,23} in response to feeding probably exerts its satiety effect²⁴ by local action on vagus afferent neurons²⁵ and involves local CCK concentrations that may be much higher than the overall concentration in the blood. This is in agreement

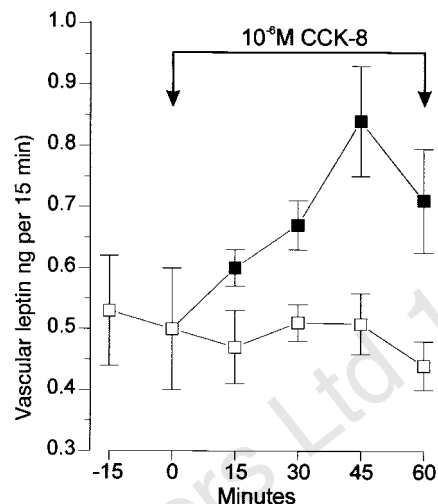


Figure 5 Kinetics of vascular leptin release from the isolated perfused rat stomach. Leptin was assayed in four experiments with CCK-8 ($1 \mu\text{M}$; filled squares). Basal leptin secretion was measured in three separate experiments (open squares). In the control stomachs, no significant change in leptin concentration was observed in the vascular effluent. Infusion of CCK-8 produced a time-dependent and substantial increase of leptin concentration. Data are means $\pm$ 1 s.e.m. and are expressed in nanograms of leptin per 15 min.

with the doses of $5\text{--}10 \mu\text{g kg}^{-1}$ that are needed to mimic this effect by intraperitoneal administration of CCK-8^{24,25}. These doses are consistent with those that have a significant effect on the release of gastric leptin, but are near threshold values with respect to CCK-induced elevation of plasma leptin, at least under our conditions. Thus, despite being an attractive hypothesis, a role for gastric leptin in CCK-induced satiety remains unproven. Alternatively, gastric leptin may have a local rather than systemic action. Leptin-sensitive vagal afferent terminals have been identified in rat stomach²⁶. On the other hand, leptin could enter the liver in quantities sufficiently large to affect glucose production and uptake as well as the hepatic expression of genes encoding key metabolic enzymes²⁷.

In summary, our findings reveal that the stomach is a source of leptin. Furthermore, the rapid mobilization of gastric leptin by feeding and exogenous CCK-8 indicates that it could be involved in CCK-mediated regulation of gastrointestinal function. □

Methods

Animals, tissues and leptin measurements. We used male Wistar, lean (*Fa/fa*) and obese (*fa/fa*) Zucker rats, which were housed in a room at 23°C with free access to food and water. They were killed by cervical dislocation and white adipose tissues of various origins were collected, together with the liver, stomach and intestine. The stomach and the intestine were opened, rinsed with saline, flattened on a glass plate with the mucosa upwards and the epithelium was scraped off and weighed. Fundic mucosa scrapings and whole epididymal fat pad were homogenized, respectively, in a KRBH solution using a glass-Teflon homogenizer. The homogenates were centrifuged at $10,000g$ for 10 min. The resulting supernatant for fundic mucosa and the infranatant for fat pad were used for leptin assays.

In another set of experiments, rats that had fasted for 18 h were injected intraperitoneally either with 0.9% saline and allowed to refeed on preweighed, standard laboratory pellet chow; or with saline (control) (1 ml per rat) or CCK-8 (Sigma Chemicals, St Louis, Missouri) diluted in 0.9% saline containing 0.1% bovine serum albumin. They were killed at appropriate times and blood was collected from the abdominal aorta, centrifuged and plasma was collected. Leptin concentrations from plasma, fundic mucosa and fat pad were estimated using a radioimmunoassay kit for mouse leptin from Linco Research Inc. (St Charles, Missouri).

PCR primers and RT-PCR conditions. Reverse transcription was done using total RNA extracted by Trizol (Gibco BRL) from fundic epithelium scrapings,

Table 1 Leptin secretion by isolated adipocytes

	Freshly isolated adipocytes	Transiently transfected adipocytes
	(Secreted leptin, ng ml^{-1} per 2 h)	
Control	0.12 ± 0.01	0.63 ± 0.04
CCK-8 ($0.1 \mu\text{M}$)	0.13 ± 0.01	0.67 ± 0.02
BRL 37344 ($0.01 \mu\text{M}$)	ND	$0.25 \pm 0.02^*$

Freshly isolated adipocytes or adipocytes overexpressing the leptin gene were incubated for 2 h as described in the Methods. Leptin secretion was measured by RIA. Values are means $\pm$ s.e.m. of four determinations in a typical experiment, representative of three experiments. ND, not determined.

* $P < 0.01$ compared with control.

white adipose tissue of various origins, and liver. First-strand cDNA was prepared from 4 µg of total RNA using murine reverse transcriptase according to the Pharmacia Biotech procedure. PCR was done using the following primers: f5'-CCTGTGGCTTTGGTCTATCTG-3' and r5'-CTGC TCAAAGC-CACCACCTCTG-3' (for the *ob* gene) and f5'-CGAGAAGATGA CCCAGAT-CATG-3' and r5'-AGTGATCTCTTCTGCATCCTG-3' (for the β -actin gene). After sample denaturation at 94 °C for 3 min, PCR was done for 40 cycles consisting of denaturation at 94 °C for 1 min, annealing at 60 °C for 1 min, and extension at 72 °C for 2 min. The amplification was terminated by a 10-min final extension step at 72 °C. In controls, reverse transcriptase was omitted. Ethidium bromide staining and 1% agarose gel electrophoresis were used to verify the size of the RT-PCR products, that is, 415 bp for the *ob* gene and 606 bp for the β -actin gene.

Antibodies. Polyclonal antibody 1 was raised in rabbit against the C-terminal fragment [137–156] of mouse leptin (Santa Cruz Biotechnology), and polyclonal antibody 2 was raised in rabbit against fragment [138–167] of mouse leptin coupled to KLH (keyhole limpet haemocyanin) using EDC (1-ethyl-3-(3-dimethyl-aminopropyl) carbodiimide hydrochloride). The two antibodies gave similar results in western blot analysis and in histological studies.

Immunoblots. Unfrozen samples of fundic scrapings and mesenteric adipose tissue were homogenized at 4 °C in a RIPA buffer containing 0.1 mg ml⁻¹ PMSF, 100 µM benzamide and 100 mM Na₃VO₄; they were then solubilized in a boiling Laemmli buffer containing β -mercaptoethanol. Proteins were separated by 12.5% polyacrylamide gel electrophoresis, transferred to nitrocellulose sheets and blotted with polyclonal antibody 1. The immune complexes were revealed by using an enhanced chemiluminescence detection system (Amersham).

Immunohistochemistry. Fed rats were injected intraperitoneally with saline (control) or with CCK-8 (300 µg kg⁻¹) and killed after 15 min. Rat stomachs were removed and fixed in Bouin's solution. Sections (4 µm thick) were incubated with antibody 1 diluted to 1 µg ml⁻¹ or with antibody 2 diluted 1:50, then with the corresponding biotinylated secondary antibody diluted 1:200, and finally in the avidin-biotin complex diluted 1:100 (Kit ABC Vectastain; Vector Laboratories). Peroxidase activity was revealed by diaminobenzidine. There was no leptin immunostaining in gastric tissues under the following conditions: (1) omission of the primary antibody; and (2) overnight preincubation of the antibodies with various concentrations of the homologous antigen, that is, 10–40 µg of leptin fragment [137–156] (antibody 1) or 10–100 µg mouse recombinant leptin (antibody 2) per millilitre of diluted antiserum.

Freshly isolated adipocytes. Adipocytes were isolated from epididymal fat pads by collagenase digestion²⁸ and incubated as a 30% adipocyte suspension in 500 µl Krebs–Ringer buffer, containing 30 mM HEPES, 1% BSA (KRBH), in the absence (controls) or presence of the various agents for 2 h. The leptin concentration in the medium was then assayed.

Transfected adipocytes. Isolated adipocytes (400 µl of a 50% adipocyte suspension) were transiently transfected by electroporation as described²⁸, with pCIS2Ob (2 µg) and pCIS2 (8 µg). Cells were diluted in 1.5 ml of Dulbecco-modified Eagle medium (DMEM) containing 0.1% BSA and 25 mM HEPES. They were incubated for 16–24 h at 37 °C in 5% CO₂/95% air, and then pooled, washed once in DMEM, resuspended (20% adipocrit) in KRBH and incubated as above for 2 h.

Isolated, vascularly perfused rat stomach. Male Wistar rats that had been deprived overnight of food were anaesthetized and laparatomized. The stomachs were isolated and vascularly perfused at 37 °C as described²⁹. Vascular effluents were collected every 15 min, stored at –20 °C and lyophilized. Samples were reconstituted in RIA buffer before leptin assay.

Received 27 March; accepted 19 June 1998.

1. Zhang, Y. *et al.* Positional cloning of the mouse *ob* gene and its human homologue. *Nature* **372**, 425–432 (1994).
2. Tartaglia, L. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 126–127 (1995).
3. Lee, G.-H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632–635 (1996).
4. Campfield, L. A. *et al.* Recombinant mouse OB protein: evidence of a peripheral signal linking adiposity and central neural networks. *Science* **269**, 546–549 (1995).
5. Pelleymounter, M., Cullen, M. & Baker, M. Effects of the *obese* gene product on body weight regulation in *ob/ob* mice. *Science* **269**, 540–543 (1995).
6. Halaas, J. L. *et al.* Weight-reducing effects of the plasma protein encoded by the *obese* gene. *Science* **269**, 543–546 (1995).

7. Masuzaki, H. *et al.* Non-adipose tissue production of leptin: leptin as a novel placenta-derived hormone in humans. *Nature Med.* **3**, 1029–1033 (1997).
8. Hainault, L., Guerre-Millo, Guichard, C. & Lavau, M. Differential regulation of adipose tissue glucose transporter in genetic obesity (fatty rat). Selective increase in the adipose cell/muscle glucose transporter (Glu4) expression. *J. Clin. Invest.* **87**, 1127–1131 (1991).
9. Dugail, I. *et al.* Adipose-tissue-specific increase in glyceraldehyde-3-phosphate dehydrogenase activity and mRNA amounts in suckling pre-obese Zucker rats. *Biochem. J.* **254**, 483–487 (1988).
10. Ahim, R. S. *et al.* Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
11. Hardie, L. J., Vernon Rayner, D., Holmes, S. & Trayhurn, P. Circulating leptin levels are modulated by fasting, cold exposure and insulin administration in lean but not Zucker (*fa/fa*) rats as measured by ELISA. *Biochem. Biophys. Res. Commun.* **223**, 660–665 (1996).
12. Weigle, D. S. *et al.* Effect of fasting, refeeding, and dietary fat restriction on plasma leptin levels. *J. Clin. Endocrinol. Metab.* **82**, 561–565 (1997).
13. Saladin, R. *et al.* Transient increase in *obese* gene expression after food intake or insulin administration. *Nature* **377**, 527–529 (1995).
14. Mac Dougald, O. A., Hwang, C. S., Fan, H. & Lane, M. D. Regulated expression of the *obese* gene product (leptin) in white adipose tissue and 3T3-L1 adipocytes. *Proc. Natl Acad. Sci. USA* **92**, 9034–9037 (1995).
15. Levy, J. R., LeGall-Salmon, E., Santos, M., Pandak, W. M. & Stevens, W. Effect of enteral versus parenteral nutrition on leptin gene expression and release into the circulation. *Biochem. Biophys. Res. Commun.* **237**, 98–102 (1997).
16. Trayhurn, P., Ducan, J. S., Rayner, D. V. & Hardie, L. J. Rapid inhibition of *ob* gene expression and circulating leptin levels in lean mice by the β_3 -adrenoceptor agonists BRL35135A and ZD2079. *Biochem. Biophys. Res. Commun.* **228**, 605–610 (1996).
17. Mantzoros, C. S. *et al.* Activation of β_3 -adrenergic receptors suppresses leptin expression and mediates a leptin-independent inhibition of food intake in mice. *Diabetes* **45**, 909–914 (1996).
18. Barrachina, M. D., Martinez, V., Wang, L., Wei, J. Y. & Taché, Y. Synergistic interaction between leptin and cholecystokinin to reduce short-term food intake in mice. *Proc. Natl Acad. Sci. USA* **94**, 10455–10460 (1997).
19. Matson, C. A., Water, M. F., Kujiper, J. L. & Weigle, D. S. Synergy between leptin and cholecystokinin (CCK) to control daily caloric intake. *Peptides* **18**, 1275–1278 (1997).
20. Koulischer, D., Moroder, L. & Deschodt-Lanckman, M. Degradation of cholecystokinin octapeptide, related fragments and analogs by human and rat plasma in vitro. *Regul. Pept.* **4**, 127–139 (1982).
21. Rose, C. *et al.* Characterization and inhibition of a cholecystokinin inactivating serine peptidase. *Nature* **380**, 403–409 (1996).
22. Buffa, R., Solcia, E. & Go, V. L. W. Immunohistochemical identification of the cholecystokinin cell in the intestinal mucosa. *Gastroenterology* **70**, 528–532 (1976).
23. Liddle, R. A., Goldfine, I. D. & Williams, J. A. Bioassay of plasma cholecystokinin in rats: effects of food, trypsin inhibitor and alcohol. *Gastroenterology* **87**, 542–549 (1984).
24. Gibbs, J., Young, C. & Smith, G. P. Cholecystokinin elicits satiety in rats with open gastric fistulas. *Nature* **245**, 323–325 (1973).
25. Smith, G. P., Jerome, C., Cushin, B. J., Eterno, T. & Simansky, K. J. Abdominal vagotomy blocks the satiety effect of cholecystokinin in the rat. *Science* **213**, 1036–1037 (1981).
26. Wang, Y. H. *et al.* Two types of leptin-responsive gastric afferent terminals: an *in vitro* single-unit study in rats. *Am. J. Physiol.* **273**, R833–R837 (1997).
27. Barzilai, N. *et al.* Leptin selectively decreases visceral adiposity and enhances insulin action. *J. Clin. Invest.* **100**, 3105–3110 (1997).
28. Tanti, J. F. *et al.* Overexpression of a constitutively active form of phosphatidylinositol 3-kinase is sufficient to promote Glut4 translocation in adipocytes. *J. Biol. Chem.* **271**, 25227–25232 (1996).
29. Bado, A. *et al.* H₃-receptor regulation of vascular gastrin and somatostatin releases by the isolated perfused rat stomach. *Yale J. Biol. Med.* **67**, 113–121 (1994).

Acknowledgements. This work was supported by INSERM and by the Ministère de l'Éducation et de la Recherche Scientifique (A.B.). L.S. was supported by MERST, K.S. by a grant from IRMAD, A.S. by a grant from FRM, and M.N.B. by a grant from Novo-Nordisk.

Correspondence and requests for materials should be addressed to M.J.M.L. (e-mail: mjmlewin@bichat.inserm.fr).

Epsin is an EH-domain-binding protein implicated in clathrin-mediated endocytosis

Hong Chen*, Silvia Fre†, Vladimir I. Slepnev*, Maria Rosaria Capua†, Kohji Takei*, Margaret H. Butler*, Pier Paolo Di Fiore‡ & Pietro De Camilli*

* Howard Hughes Medical Institute and Department of Cell Biology, Yale University School of Medicine, 295 Congress Avenue, New Haven, Connecticut 06510, USA

† Department of Experimental Oncology, European Institute of Oncology, Milan 20141, Italy

‡ Istituto di Microbiologia, Università di Bari, Bari 70124, Italy

During endocytosis, clathrin and the clathrin adaptor protein AP-2 (ref. 1), assisted by a variety of accessory factors, help to generate an invaginated bud at the cell membrane^{2,3}. One of these factors is Eps15, a clathrin-coat-associated protein that binds the α -adaptin subunit of AP-2 (refs 4–8). Here we investigate the function of Eps15 by characterizing an important binding partner for its

region containing EH domains⁹; this protein, epsin, is closely related to the *Xenopus* mitotic phosphoprotein MP90 (ref. 10) and has a ubiquitous tissue distribution. It is concentrated together with Eps15 in presynaptic nerve terminals, which are sites specialized for the clathrin-mediated endocytosis of synaptic vesicles. The central region of epsin binds AP-2 and its carboxy-terminal region binds Eps15. Epsin is associated with clathrin coats *in situ*, can be co-precipitated with AP-2 and Eps15 from brain extracts, but does not co-purify with clathrin coat components in a clathrin-coated vesicle fraction. When epsin function is disrupted, clathrin-mediated endocytosis is blocked. We propose that epsin may participate, together with Eps15, in the molecular rearrangement of the clathrin coats that are required for coated-pit invagination and vesicle fission.

The core of the binding consensus of the EH domains of Eps15 is the sequence asparagine-proline-phenylalanine (NPF)¹¹. This sequence is present in non-neuronal isoforms of proteins of the synaptojanin and AP180/CALM families, which have been implicated in clathrin-mediated endocytosis^{12–14}. Members of these families interact with the EH domains of Eps15 and of the yeast protein Pan1^{15,16}. Clathrin-mediated endocytosis of synaptic vesicles probably represents a specialization of the clathrin-mediated endocytosis typical of all cells², but the nerve-terminal isoforms of AP180 and synaptojanin do not contain NPF motifs^{12,13,15,17}, raising the possibility that Eps15 may be selectively involved in clathrin-mediated internalization of receptors and not in synaptic vesicle recycling. We therefore tested whether Eps15, and the homologous protein Eps15R (refs 6, 18), are concentrated in nerve endings, which would be expected for proteins that participate in synaptic vesicle endocytosis.

Immunofluorescence staining of frozen sections revealed differential expression of Eps15 and Eps15R in various rat brain regions, but produced in all regions a predominant nerve-terminal pattern of immunoreactivity, similar to that of synaptophysin, a synaptic-vesicle membrane protein^{6,19} (Fig. 1a, and results not shown). There was an enrichment of Eps15 and Eps15R in brain relative to other tissues (Fig. 1b), consistent with an important role of clathrin-dependent synaptic vesicle recycling in this tissue.

We searched for nerve-terminal binding partners for the EH-domain-containing regions of Eps15 and Eps15R by reacting fusion

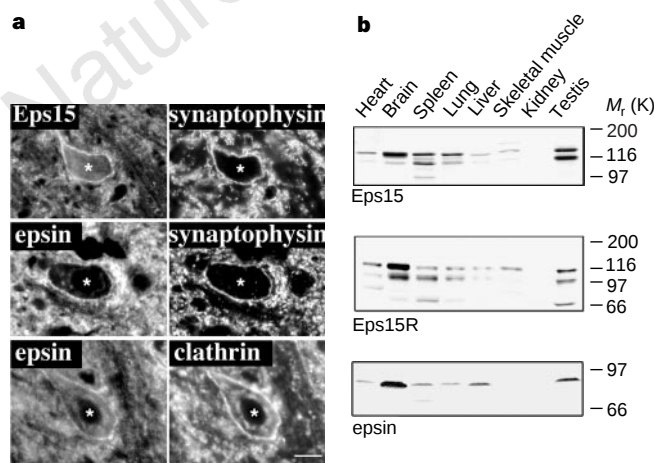


Figure 1 Eps15 and epsin are highly expressed in the brain and are concentrated in nerve terminals. **a**, Double-immunofluorescence micrographs of rat brain frozen sections showing the co-localization of Eps15 and epsin with synaptophysin and clathrin in nerve terminals. Epsin and clathrin are also localized in the Golgi complex, visible as perinuclear immunostaining. Asterisks, neuronal perikarya. Scale bar, 20 μ m. **b**, Tissue distribution of Eps15, Eps15R and epsin, as determined by western blotting of total homogenates of rat tissues. The multiple bands stained by both anti-Eps15 and anti-Eps15R probably represent alternatively spliced isoforms of these two proteins¹⁸.

proteins comprising glutathione-S-transferase (GST) and these regions (GST-EH_{eps15} and GST-EH_{eps15R}, respectively) with rat brain cytosol in a gel overlay assay. The two fusion proteins, but not GST alone (Fig. 2a), specifically recognized one predominant band corresponding to a relative molecular mass of ~94K (referred to here as epsin, for Eps15-interacting protein) and minor bands of smaller molecular masses. A protein band of 94K (p94) was identified previously by affinity chromatography as a binding partner for the appendage domain of α -adaptin, a subunit of the clathrin adaptor AP-2 (ref. 20), but was not characterized further. As shown by Fig. 2, antibodies raised against one of the three reported peptide sequences of bovine p94 specifically recognized in rat brain a protein band that co-migrated with epsin in SDS-PAGE (Fig. 2a) and which could be affinity-purified on immobilized GST-EH_{eps15} (Fig. 2b). The same antibodies immunodepleted epsin from rat brain cytosol (Fig. 2c).

Degenerate oligonucleotides designed from the reported peptide sequences of bovine p94 were used to clone by polymerase chain reaction (PCR) a fragment of rat epsin that was then used to screen a λ ZAPII rat brain complementary DNA library (from Stratagene). Clone 16 (accession no. AF018261) included a full-length 576-amino-acid open reading frame (Fig. 3) which contained stretches identical to the three peptides of bovine p94, except that there was an arginine-to-lysine substitution at residue 128. Searches of the databases revealed several related complete and partial sequences, including the yeast sequences YDL161w and YLR206w. The more closely related full-length sequence is that of MP90, a *Xenopus* mitotic phosphoprotein¹⁰. The domain comprising the first 165 amino acids of epsin is the most conserved phylogenetically and is present in proteins otherwise very divergent from epsin; it appears therefore to define a new protein module. The central 150-amino-acid region is strikingly characterized by the presence of several repeats of the triplet aspartate-proline-tryptophan (DPW), which are almost completely conserved from *Xenopus* to rat. The Cdc2 phosphorylation site of MP90 is localized in this region¹⁰ and is conserved in epsin. The C-terminal region of epsin contains three NPF repeats (Fig. 3), consistent with the identification of epsin as a partner for the EH domains of Eps15 (ref. 11).

By northern blotting, a 2.6-kilobase (kb) epsin-specific band was labelled in all of several tissues tested, suggesting a housekeeping role for this protein; epsin immunoreactivity was also detectable in all tissues by western blotting. Like Eps15 and Eps15R, the highest

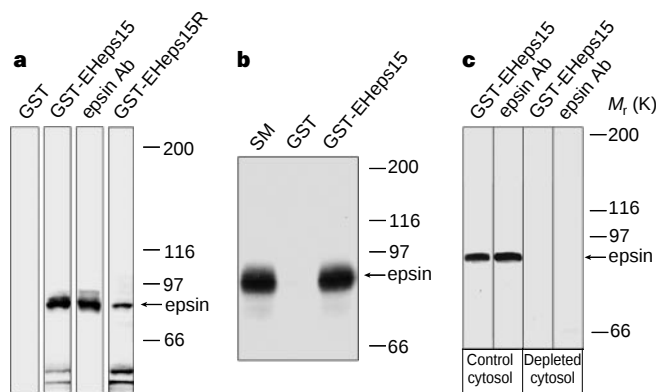


Figure 2 EH-domain-containing regions of Eps15 and Eps15R bind to a 94K protein (epsin) in rat brain. **a**, SDS-PAGE gels of rat brain cytosol were overlaid with GST fusion proteins comprising the three EH domains of Eps15 or Eps15R (GST-EH_{eps15} and GST-EH_{eps15R}) or antibodies raised against the p94 protein (ref. 20; epsin Ab). **b**, Affinity purification of rat brain Triton X-100 extract on either GST or GST-EH_{eps15}. The starting material (SM) and material bound to the beads were labelled by western blotting with antibodies directed against p94 (epsin). **c**, GST-EH_{eps15} overlay or anti-epsin western blotting of rat brain cytosol before or after immunodepletion with anti-p94 antibodies.

a

```

1  MSTSSLRQMKNIIVHNYSEAEIKVREATSNDPWGPSSSLMSEIADLTYNV
51  VAFSEIMSIWKRLNDHGKNWRHVYKAMTLMEYLKGTGSEVSQCKENM
101 YAVQTLKDFQYVDRDQDQGVNVREKAKQLVALLRDEDRLEERAHAKT
151 KEKLAQTATASSAAVGSGPPPEAEQAWPQSSGEEELQLQLALAMSKEEAD
201 QPPSCGPEDDVQLQLALSLSREEHDKEERIRRGDDLRLQMAIEESKRETG
251 GKEESSLMDLADVFTTPALPQASDPWGGPASVPTAVPVAAAADPWGAPA
301 VPPAADPWGGAAPTASDPWRPAAAPTGPSVDPWGGTPAPAAAGGPTSDP
351 WGSADGGAPVSGPPSSDPWAPAPAFSDDPWGGSPAKPSSNGTAAVGGFDTE
401 PDEFSDFDRLRTALPTSGSSTGELELLAGEVPARSPGAFMSGVGGSLAE
451 SVGSPPPAATPTHTPTTRKTPESFLGPNAAALVDLSVSRPGPTPPGAKA
501 SNPELPSGAPATGPSVTNPEQAPPATLTNLRLSPVPPVPGAPPTYIS
551 PLGGGPGLPMPMPGPPAPNTNPELL
  
```

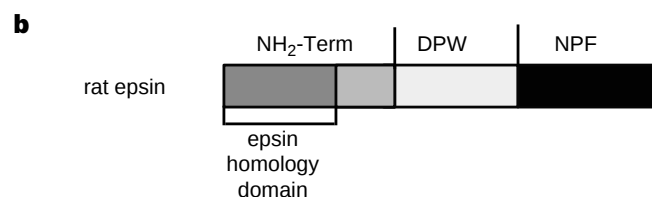


Figure 3 Primary sequence of rat epsin. **a**, Deduced amino-acid sequence of rat epsin. The three peptides sequences of bovine p94 obtained by Edman degradation²⁰ are underlined with thick grey bars. DPW sequences are in bold and NPF motifs are underlined. The putative Cdc2 phosphorylation site is indicated by an asterisk. **b**, Domain structure of epsin.

concentration of epsin was found in the brain (Fig. 1b). Epsin was concentrated in nerve terminals, where it co-localized with synaptophysin and with clathrin. It was also present, together with clathrin, in the Golgi complex (Fig. 1a). In subcellular fractions of rat brain, epsin and Eps15 were co-enriched in synaptosomes (P2 fraction) and had a roughly parallel distribution in other fractions (Fig. 4a). Epsin, like Eps15²¹, was present but not enriched, in a clathrin-coated vesicle fraction prepared according to ref. 22, as was also found for other accessory proteins of clathrin-mediated endocytosis (amphiphysin I, dynamin I, and synaptojanin I; refs 23, 24 and 12, respectively). In contrast, intrinsic components of the clathrin coat (clathrin, AP-2 subunits and AP180)² were enriched in this fraction (Fig. 4a). Immunogold labelling of synaptic membranes confirmed a specific association of epsin with clathrin coats (Fig. 4b).

These results were supported by analysis of the subcellular distribution of epsin in transfected CHO cells briefly permeabilized before fixation. In cells with low-to-moderate expression, epsin immunofluorescence had a punctate pattern that was very similar to that of clathrin, AP-2 and Eps15 (Fig. 4c), and epsin immunogold labelled primarily clathrin-coated pits (Fig. 4b). In cells with high expression, there was additional labelling of the cytoplasmic face of the plasmalemma (not shown), indicating that epsin may interact with other molecules of the cortical cytoplasm besides clathrin-coated components.

To confirm the binding of epsin to α -adaptin and Eps15, and to identify binding regions for these proteins, brain cytosol was affinity-purified on GST fusion proteins of the three main domains of epsin. The NPF domain bound Eps15, as expected, whereas the DPW domain bound α -adaptin (both α_a and α_c subunits) (Fig. 5a).

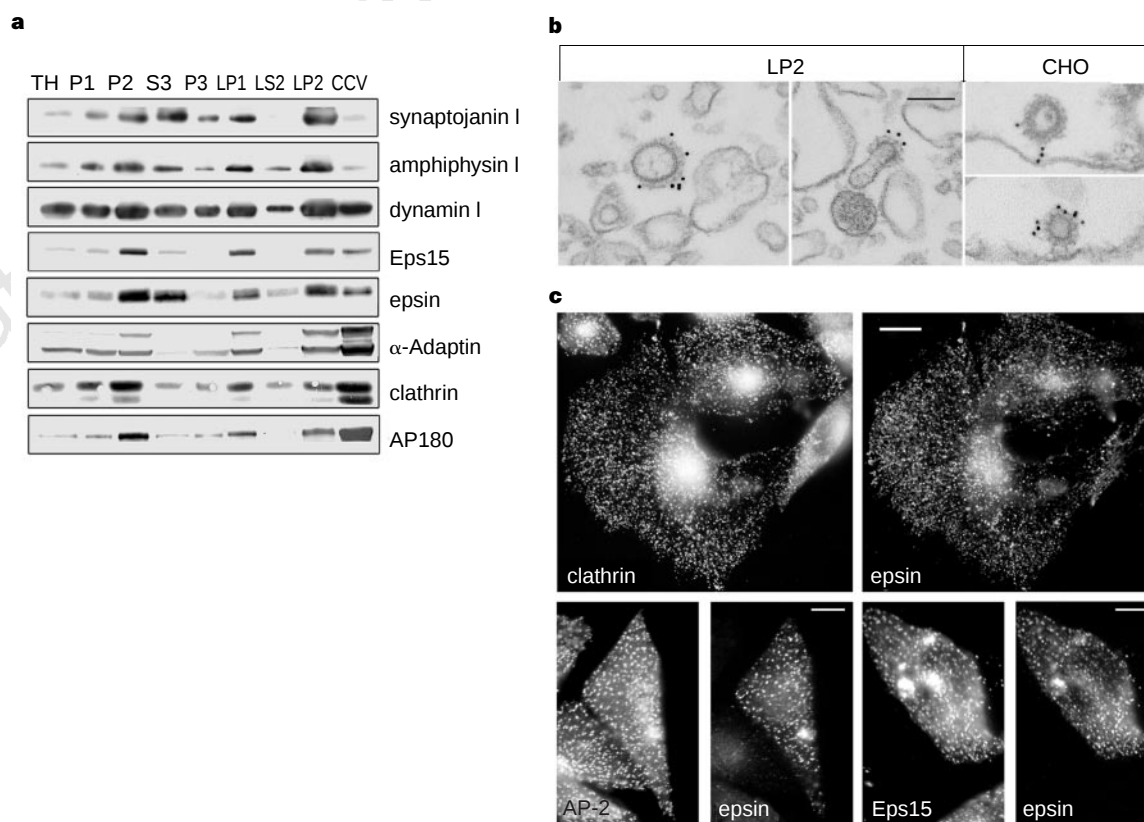


Figure 4 Subcellular localization of epsin. **a**, Distribution of epsin and other endocytic proteins in subcellular fractions of rat brain (TH to LP2) obtained according to ref. 28. A clathrin-coated vesicle fraction (CCV) was prepared as described²². **b**, Electron microscopy of immunogold labelling for epsin in an LP2 fraction of rat brain after incubation with brain cytosol, ATP and GTP- γ S, and in

transiently transfected CHO cells briefly lysed before fixation. Scale bar, 100 nm. **c**, Immunofluorescence of CHO cells transiently transfected with epsin cDNA and briefly permeabilized before fixation showing the co-localization of epsin with other endocytic proteins. Scale bar, 15 μ m.

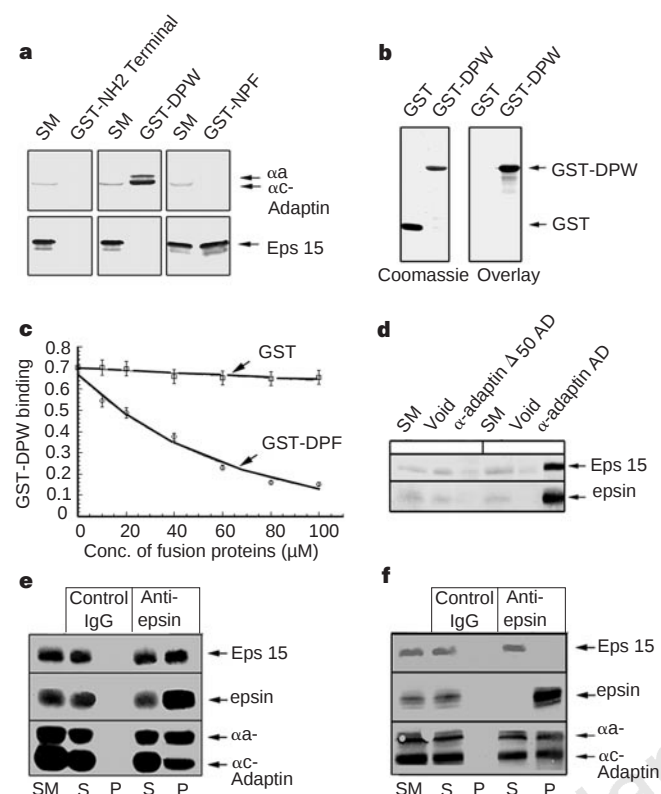


Figure 5 Interaction of epsin with α -adaptin and Eps15. **a**, A rat brain Triton X-100 extract was affinity-purified on GST fusion proteins comprising the three portions of epsin (see Fig. 3b). Starting material and bound material were probed by western blotting for α -adaptin or Eps15. **b**, Coomassie-blue staining and α -adaptin (appendage domain) overlay of GST and a GST-DPW fusion protein. **c**, ELISA assay showing that the DPW domain of Eps15 competes with the DPW domain of epsin for binding to the immobilized appendage domain of α -adaptin. **d**, Affinity purification of a rat brain Triton X-100 extract onto α -adaptin appendage domain (α -adaptin AD)²⁰ or a truncated appendage domain with a 50-amino-acid N-terminal deletion (α -adaptin Δ 50 AD). Void, unbound material. **e, f**, Western blot analysis of immunoprecipitates from a Triton X-100 extract of a total homogenate of rat brain (**e**) or brain cytosol (**f**) with anti-epsin antibodies or control IgGs. SM, starting material; S, supernatant; P, pellet.

Further analysis by gel overlay revealed that there was direct interaction not only of the NPF domain with Eps15 (not shown, and Fig. 2a), but also of the DPW domain with the appendage domain of α -adaptin (Fig. 5b). As shown by an enzyme-linked immunosorbent assay (ELISA), the latter interaction was competitively inhibited by the α -adaptin binding domain of Eps15 (the so-called DPW domain)^{4,7} (Fig. 5c, and results not shown). Consistent with a binding of epsin and Eps15 to the same region of α -adaptin, a 50-amino-acid N-terminal truncation of the appendage domain of α -adaptin abolished binding of both Eps15⁴ and epsin (Fig. 5d). The occurrence of these interactions *in vivo* was confirmed by co-precipitation experiments, in which antibodies directed against epsin co-precipitated both α -adaptin and Eps15 from a Triton X-100 extract of a total brain homogenate (Fig. 5e). However, the same antibodies co-precipitated only α -adaptin from the cytosol (Fig. 5f), suggesting that the Eps15-epsin interaction occurs only at the membrane.

If epsin acts as a link protein between AP-2 and EH-domain-containing proteins, the overexpression of its AP-2-binding domain may have a dominant-negative effect on clathrin-mediated endocytosis. As shown by Fig. 6a–d, CY3-transferrin uptake²⁵ by CHO cells is inhibited by expression of the DPW domain of epsin but not by expression of its N-terminal domain. Likewise, microinjection of a GST-DPW construct in CV-1 cells blocked the clathrin-

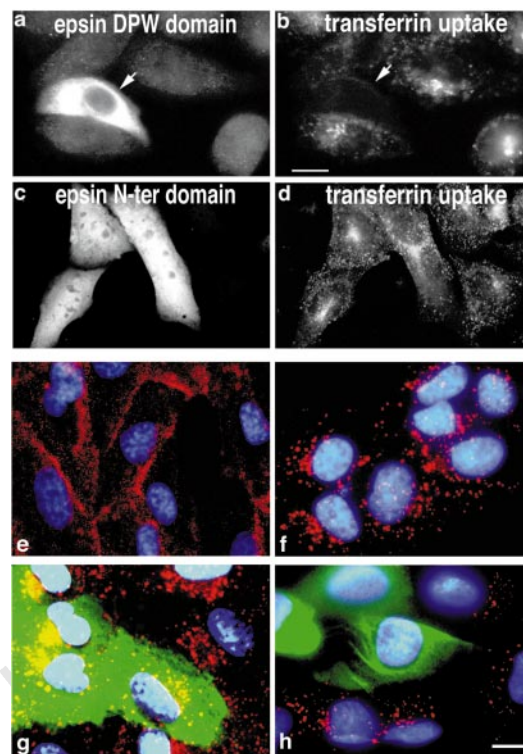


Figure 6 Perturbation of epsin function impairs receptor-mediated endocytosis. **a–d**, CHO cells were transiently transfected with the DPW domain (**a** and **b**, arrow) or the amino-terminal domain (**c**, **d**) of epsin and exposed to CY3-transferrin for 20 min before fixation and immunolabelling for the transfected proteins. Scale bar, 14 μ m. **e–h**, Colour confocal micrographs demonstrating that microinjection of CV-1 cells with anti-epsin antibodies blocks EGF-receptor internalization. Cells were exposed to rhodamine-EGF for 1 h at 4 °C (**e**, note cell-surface-bound EGF) and then incubated for 15 min at 37 °C, briefly acid-washed and fixed (**f–h**). Cells in **g** and **h** were injected with control IgGs (**g**) or affinity-purified anti-epsin antibodies (**h**), respectively. Red, EGF; green, injected antibodies; blue, DAPI staining. Scale bar, 15 μ m. Results are typical and representative of four experiments in which at least 100 cells were injected per experiment.

dependent internalization of receptors for epidermal growth factor (EGF) (results not shown). We also tested the effect of epsin disruption on clathrin-mediated endocytosis by injecting antibody into CV-1 cells. Injection of affinity-purified antibodies directed against epsin, but not of control antibodies, inhibited the internalization of both EGF (Fig. 6e–h) and transferrin (results not shown).

Collectively, our results indicated that epsin participates in clathrin-dependent endocytosis, including endocytosis at the synapse. Epsin, like Eps15 (ref. 21), is clearly not a major intrinsic component of the clathrin coat. As proposed for Eps15 (refs 8, 21), it may assist clathrin during oligomerization and during the progressive rearrangement of the clathrin lattice that leads to invagination and fission. As epsin and Eps15 can bind to each other at the membrane, they could form dynamic molecular bridges between neighbouring AP-2 complexes. This effect may be amplified by the property of Eps15, an elongated molecule, to form dimers and tetramers²⁶. An entry in the GenBank database of a partial sequence that is likely to represent mouse epsin (AF057285) reveals that epsin binds intersectin (AF032118), another EH-domain-containing protein. Intersectin is homologous with DAP 160, a *Drosophila* dynamin-interacting protein²⁷. Thus, epsin seems to participate in a network of protein–protein interactions that assists clathrin-mediated endocytosis and which may have additional functions in the cortical cytoplasm. As epsin is also localized in the Golgi

complex area, EH-domain-mediated interactions of epsin may have a general role in clathrin-mediated budding. □

Methods

Antibodies. Rabbit antibodies directed against epsin and Eps15 were obtained using the following immunogens: bovine epsin peptide ARQLVALLRDEDLREE²⁰, GST fusion proteins of the DPW and NPF domains of epsin, and of the EH and DPF domains of Eps15. Antibodies directed against anti-Eps15R, dynamin I, amphiphysin I and synaptotagmin I have been described^{15,18,23}. A monoclonal antibody directed against the appendage domain of α -adaptin was from Sigma; antibodies directed against clathrin light chain, synaptophysin and synaptotagmin were gifts from R. Jahn.

Cloning and sequencing of epsin cDNA. Degenerate primers corresponding to two peptide sequences of bovine p94 (ref. 20; NIVHNY and FQYVDR) were used in PCR reactions with a rat brain cDNA library as template. A 300-bp PCR product was amplified, subcloned into a TA cloning vector (Invitrogen), radiolabelled and used to screen a rat brain cDNA library (Stratagene).

Production and affinity purification of fusion proteins. Portions of the coding region of rat epsin encoding amino-acid residues 1–248 (N-terminal domain), 249–401 (DPW domain), and 402–576 (NPF domain), were amplified by PCR using Vent polymerase (NEB, Beverly, MA). PCR fragments were digested by either *Eco*R1 and *Sal*I, or *Eco*R1 and *Xho*I, and subcloned into the pGEX4T series of vectors (Pharmacia). GST fusion proteins were produced and purified on a glutathione–Sepharose 4B column (Pharmacia) according to the manufacturer's protocol.

Cell transfection. Full-length epsin cDNA was subcloned into pcDNA3.1/His B (Invitrogen). Epsin fragments were subcloned into pcDNA3 with an N-terminal Flag tag. These constructs were transfected in CHO cells by the lipofectamine method (GIBCO) and cells were analysed 18–24 h after transfection. For CY3-transferrin (Molecular Probes) uptake²⁵, cells were incubated in the presence of the probe for 20 min at room temperature before fixation in 4% paraformaldehyde/120 mM sodium phosphate buffer, pH 7.4.

Cell microinjections. CV-1 cells were plated at a density of 5×10^4 cells on round glass gelatin-coated coverslips. Injection pressure was set at 30–80 hPa and the injection time at 0.3–0.5 s. Affinity-purified anti-epsin polyclonal antibodies or rabbit IgG (Jackson ImmunoResearch) were injected at 1.2 mg ml^{-1} . For EGF internalization assay, microinjected cells were incubated with $1 \mu\text{g ml}^{-1}$ rhodamine–EGF (Molecular Probes) in DMEM medium at 4°C for 1 h. The EGF-containing medium was then replaced with warm DMEM and cells were further incubated at 37°C for 15 min, followed in some cases by an acid wash before fixation in order to remove cell-surface EGF. For transferrin internalization assay, microinjected cells were incubated with $50 \mu\text{g ml}^{-1}$ rhodamine–transferrin (Molecular Probes) in DMEM medium at 37°C for 60 min. Following internalization, cells were fixed with 4% paraformaldehyde, and immunostained with FITC-conjugated goat anti-rabbit secondary antibodies (Jackson ImmunoResearch) to identify injected cells. Nuclear counterstaining was done by incubating coverslips for 5 min at 25°C with DAPI (Sigma).

Immunoprecipitation and immunodepletion. Rat brain cytosol or a Triton X-100 extract of a total brain homogenate (10 mg ml^{-1}) were incubated for 2 h at 4°C with affinity-purified anti-epsin IgGs or control IgG bound to protein A–Sepharose beads. Immunoprecipitates were washed three times for 5 min with TBST buffer (20 mM Tris, pH 7.4, 150 mM NaCl, 1% Triton X-100, 1 mM PMSF), and analysed by SDS–PAGE and western blotting. For immunodepletion of the cytosol, a large excess of protein A–Sepharose precoated with IgGs directed against the bovine p94 peptide was added to the cytosol, incubated for 4 h at 4°C and then separated by centrifugation.

Electron microscopy. Immunogold labelling of synaptic membrane (LP2 fractions)²⁸ was performed as described²⁴ after incubation of these membranes in the presence of brain cytosol, ATP and GTP- γ S, to increase the number of clathrin-coated endocytic intermediates²⁴. Transfected CHO cells were mechanically disrupted in cytosolic buffer by a nitrocellulose paper 'rip-off' method, then fixed in 3% formaldehyde and 0.1% glutaraldehyde, processed for immunogold labelling²⁴ while still attached on the Petri dish, and finally scraped and epon-embedded for ultrathin sectioning.

ELISA assay. 96-well microtitre plates were coated with α -adaptin appendage domain by overnight incubation at 4°C with $20 \mu\text{g ml}^{-1}$ appendage domain in 100 mM borate buffer, pH 8.5, followed by blocking with 5% BSA. $2 \mu\text{M}$

biotinylated DPW domain was then added to each well in the presence of GST or GST fused to the DPF domain of Eps15, and incubated for 2 h at room temperature, followed by 1 h incubation with alkaline-phosphatase-conjugated streptavidin. Colorimetric reactions were quantified by a microplate reader.

Miscellaneous procedures. Rat brain homogenate was subcellular-fractionated as described²⁸. SDS–PAGE, western blotting, gel overlays, protein assays and immunofluorescence were done as described²⁹. For Fig. 4c, cells were exposed to 0.03% saponin in cytosolic buffer²⁴ for a minute before fixation to elute soluble proteins. Preparation of Triton X-100 rat brain extracts and affinity-purification methods from these extracts on immobilized GST fusion proteins were carried out as described³⁰.

Received 21 April; accepted 20 June 1998.

- Robinson, M. S. The role of clathrin, adaptors and dynamin in endocytosis. *Curr. Opin. Cell Biol.* **6**, 538–544 (1994).
- Cremona, O. & De Camilla, P. Synaptic vesicle endocytosis. *Curr. Opin. Neurobiol.* **7**, 323–330 (1997).
- Kirchhausen, T., Bonifacino, J. S. & Riezman, H. Linking cargo to vesicle formation: receptor tail interactions with coat proteins. *Curr. Opin. Cell Biol.* **9**, 488–495 (1997).
- Benmerah, A., Begue, B., Dautry-Varsat, A. & Cerf-Bennussan, N. The ear of α -adaptin interacts with the COOH-terminal domain of the Eps 15 protein. *J. Biol. Chem.* **271**, 12111–12116 (1996).
- Benmerah, A. *et al.* AP-2/Eps15 interactions is required for receptor-mediated endocytosis. *J. Cell Biol.* **140**, 1055–1062 (1998).
- Carbone, R. *et al.* eps15 and eps15R are essential components of the endocytic pathway. *Cancer Res.* **57**, 5498–5504 (1997).
- Iannolo, G. *et al.* Mapping of the molecular determinants involved in the interaction between eps15 and AP-2. *Cancer Res.* **57**, 240–245 (1997).
- Tebar, F., Sorkina, T., Sorkin, A., Ericsson, M. & Kirchhausen, T. Eps15 is a component of clathrin-coated pits and vesicles and is located at the rim of coated pits. *J. Biol. Chem.* **271**, 28727–28730 (1996).
- Wong, W. T. *et al.* A protein-binding domain, EH, identified in the receptor tyrosine kinase substrate Eps15 and conserved in evolution. *Proc. Natl Acad. Sci. USA* **92**, 9530–9534 (1995).
- Stukenberg, P. T. *et al.* Systematic identification of mitotic phosphoproteins. *Curr. Biol.* **7**, 338–348 (1997).
- Salcini, A. E. *et al.* Binding specificity and *in vivo* targets of the EH domain, a novel protein–protein interaction module. *Genes Dev.* **11**, 2239–2249 (1997).
- McPherson, P. S. *et al.* A presynaptic inositol-5-phosphatase. *Nature* **379**, 353–357 (1996).
- Dreyling, M. H. *et al.* The t(10;11)(p13;q14) in the U937 cell line results in the fusion of the AF10 gene and CALM, encoding a new member of the AP-3 clathrin assembly protein family. *Proc. Natl Acad. Sci. USA* **93**, 4804–4809 (1996).
- Srinivasan, S. *et al.* Disruption of three phosphatidylinositol-polyphosphate 5-phosphatase genes from *Saccharomyces cerevisiae* results in pleiotropic abnormalities of vacuole morphology, cell shape, and osmohomeostasis. *Eur. J. Cell Biol.* **74**, 350–360 (1997).
- Haffner, C. *et al.* Synaptotagmin I: localization on coated endocytic intermediates in nerve terminals and interaction of its 170 kDa isoform with Eps15. *FEBS Lett.* **419**, 175–180 (1997).
- Wendland, B. & Emr, S. D. Pan1p, Yeast eps15, Functions as a multivalent adaptor that coordinates protein–protein interactions essential for endocytosis. *J. Cell Biol.* **141**, 71–84 (1998).
- Zhou, S., Sousa, R., Tannery, N. H. & Lafer, E. M. Characterization of a novel synapse-specific protein. *J. Neurosci.* **12**, 2144–2155 (1992).
- Coda, L. *et al.* Eps15R is a tyrosine kinase substrate with characteristics of a docking protein possibly involved in coated pits-mediated internalization. *J. Biol. Chem.* **273**, 3003–3012 (1998).
- Navone, F. *et al.* Protein p38: an integral membrane protein specific for small vesicles of neurons and neuroendocrine cells. *J. Cell Biol.* **103**, 2511–2527 (1986).
- Wang, L. H., Sudhof, T. C. & Anderson, R. G. The appendage domain of α -adaptin is a high affinity binding site for dynamin. *J. Biol. Chem.* **270**, 10079–10083 (1995).
- Cupers, P., Jadhav, A. P. & Kirchhausen, T. Assembly of clathrin coats disrupts the association between Eps15 and AP-2 adaptors. *J. Biol. Chem.* **273**, 1847–1850 (1998).
- Maycox, P. R., Link, E., Reetz, A., Morris, S. A. & Jahn, R. Clathrin-coated vesicles in nervous tissue are involved primarily in synaptic vesicle recycling. *J. Cell Biol.* **118**, 1379–1388 (1992).
- Bauerfeind, R., Takei, K. & De Camilli, P. Amphiphysin I is associated with coated endocytic intermediates and undergoes stimulation-dependent dephosphorylation in nerve terminals. *J. Biol. Chem.* **272**, 30984–30992 (1997).
- Takei, K., McPherson, P. S., Schmid, S. L. & De Camilli, P. Tubular membrane invaginations coated by dynamin rings are induced by GTP- γ S in nerve terminals. *Nature* **374**, 186–190 (1995).
- Cameron, P. L., Sudhof, T. C., Jahn, R. & De Camilli, P. Localization of synaptophysin with transferrin receptors: implications for synaptic vesicle biogenesis. *J. Cell Biol.* **115**, 151–164 (1991).
- Cupers, P., ter Haar, E., Boll, W. & Kirchhausen, T. Parallel dimers and antiparallel tetramers formed by epidermal growth factor receptor pathway substrate clone 15. *J. Biol. Chem.* **272**, 33430–33434 (1997).
- Roos, J. & Kelly, R. B. Dap160, a neural-specific EH- and multiple SH3-domain containing protein that interacts with Drosophila dynamin. *J. Biol. Chem.* (in the press).
- Huttner, W. B., Schiebler, W., Greengard, P. & De Camilli, P. Synapsin I (protein I), a nerve terminal-specific phosphoprotein. *J. Cell Biol.* **96**, 1374–1388 (1983).
- David, C., McPherson, P. S., Mundigl, O. & De Camilli, P. A role of amphiphysin in synaptic vesicle endocytosis suggested by its binding to dynamin in nerve terminals. *Proc. Natl Acad. Sci. USA* **93**, 331–335 (1996).
- Ringstad, N., Nemoto, Y. & De Camilli, P. The SH3p4/SH3p8/SH3p13 protein family: binding partners for synaptotagmin and dynamin via a Grb2-like Src homology 3 domain. *Proc. Natl Acad. Sci. USA* **94**, 8569–8574 (1997).

Acknowledgements. We thank Y. Nemoto for help in preliminary experiments, R. Bauerfeind and O. Cremona for advice, L. Daniell and M. Salazar for help with electron microscopy, X. Zhang for technical assistance and support, and M. M. Zhou for discussion. This work was supported in part by grants from the NIH, the HFSP and the US Army Medical Research and Development Command (to P.D.C.), and from the Associazione Italiana Ricerca sul Cancro, the Consiglio Nazionale delle Ricerche, the European Community (BIOMED-2 Programme), the Armenise-Harvard Foundation and the Ferrero Foundation (to P.D.C.).

Correspondence and requests for materials should be addressed to P.D.C. (e-mail: pietro.decamilli@yale.edu). The accession number for the rat epsin nucleotide and amino-acid sequences is AF018261 in GenBank.

GPI-anchored proteins are organized in submicron domains at the cell surface

Rajat Varma & Satyajit Mayor

National Centre for Biological Sciences, TIFR Centre, IISc Campus, Post Box 1234, Bangalore 560012, India

Lateral heterogeneities in the classical fluid-mosaic model of cell membranes are now envisaged as domains or 'rafts' that are enriched in (glyco)sphingolipids, cholesterol, specific membrane proteins and glycosylphosphatidylinositol (GPI)-anchored proteins¹. These rafts dictate the sorting of associated proteins and/or provide sites for assembling cytoplasmic signalling molecules². However, there is no direct evidence that rafts exist in living cells^{3,4}. We have now measured the extent of energy transfer between isoforms of the folate receptor bound to a fluorescent analogue of folic acid, in terms of the dependence of fluorescence polarization on fluorophore densities in membranes⁵. We find that the extent of energy transfer for the GPI-anchored folate-receptor isoform is density-independent, which is characteristic of organization in sub-pixel-sized domains at the surface of living cells; however, the extent of energy transfer for the transmembrane-anchored folate-receptor isoform was density-dependent, which is consistent with a random distribution. These domains are likely to be less than 70 nm in diameter and are disrupted by removal of cellular cholesterol. These results indicate that lipid-linked proteins are organized in cholesterol-dependent submicron-sized domains. Our methodology offers a new way of monitoring nanometre-scale association between molecules in living cells.

Although the existence of phase-separated regions has been demonstrated in artificial membranes by using diverse biophysical techniques^{3,6,7}, the principal evidence for rafts in cells rests on the observation that specific membrane proteins, including GPI-anchored proteins and lipid-linked non-receptor tyrosine kinases, (glyco)sphingolipids and cholesterol form detergent-insoluble complexes in cold Triton X-100 (ref. 2). Co-purification with such complexes is often taken as evidence for the presence of domains in membranes, but this alone cannot constitute proof of pre-existing domains in living cells^{4,8,9}.

GPI-anchored proteins are a diverse set of (glyco)lipid-linked cell-surface, exoplasmic eukaryotic proteins¹⁰ which appear to be diffusely distributed at the cell surface, thus belying the existence of rafts^{9,11–13}. Paradoxically, it is the functional properties of GPI-anchored proteins that provide the strongest case for lateral segregation of these proteins in membranes^{2,4,14}. The failure to observe GPI-anchored protein domains may be because of their mobility in the plane of the membrane, coupled with the limited resolution of the fluorescence microscope (>300 nm), and the inability to detect small aggregates of lipid species by current electron microscopic techniques¹⁵. Analysis of the diffusion patterns of several molecules at the cell surface, including GPI-anchored proteins, transmembrane anchored proteins and a ganglioside, has shown that these proteins are transiently confined to small domains (200–300 nm) in relatively undifferentiated areas of the membrane, independent of the nature of their membrane anchor^{1,16,17}. Therefore if these domains or rafts containing GPI-anchored proteins exist, they must be smaller than 300 nm, dynamic, and contain few proteins.

To demonstrate the existence of domains, we have measured the distance between GPI-anchored proteins and determined whether they are organized non-randomly, by using fluorescence resonance

energy transfer (FRET) microscopy. FRET has been used before to show that a GPI-anchored viral glycoprotein and 5' nucleotidase were not significantly clustered at steady state, whereas a ganglioside, GM1, was clustered^{18–20}.

FRET is accompanied by a loss in polarization of the net emission because there is almost total loss of polarization during energy transfer⁵, so loss of fluorescence anisotropy provides a sensitive assay for FRET between like fluorophores²¹. Figure 1 shows two possible arrangements of fluorophores as visualized in a microscope pixel focused at the cell surface, and the predicted change in anisotropy consistent with the given geometry and pixel fluorescence intensity.

To compare the organization of GPI-anchored proteins with a transmembrane protein, we obtained stable clones of Chinese hamster ovary (CHO) cells transfected with a GPI-anchored folate receptor (FR-GPI) or a chimaeric transmembrane-anchored folate receptor (FR-TM). We used the folate-binding domain as a reporter molecule as it is a monomeric protein and may be exogenously labelled by binding monovalently to a fluorescent folic acid analogue, *N*^α-pteroyl-*N*^ε-(4'-fluorescein-thiocarbonyl)-L-lysine (PLF)⁹.

We found that fluorescence anisotropy values at the surface of PLF-labelled, FR-GPI-expressing cells were constant over the entire fluorescence intensity range (Fig. 2a–d), whereas anisotropy values were inversely dependent on fluorescence intensity for the transmembrane-linked protein FR-TM (Fig. 2e–h). The constant anisotropy value for different fluorescence intensities seen for FR-GPI is consistent with the possibility that there is no energy transfer taking place between the FR-GPI species, or that GPI-anchored proteins are organized in small (sub-pixel resolution) domains at the cell

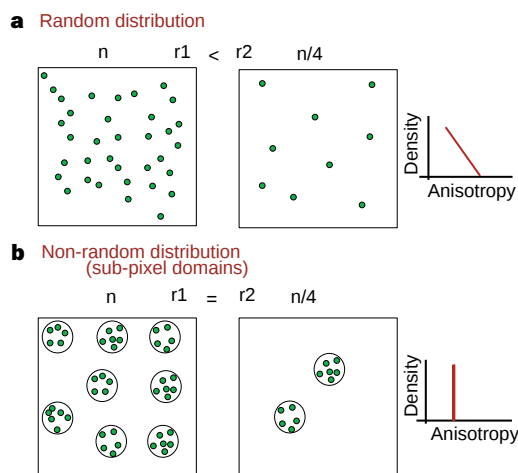


Figure 1 Possible outcomes of the fluorescence depolarization experiment for different arrangements of fluorescently labelled GPI-anchored proteins in a microscope pixel. If GPI-anchored proteins are present as uniform random distributions in membranes (**a**), the average distance between individual protein molecules on a two-dimensional surface increases by a factor of two upon a fourfold decrease in fluorophore density. As the fluorescence intensity measured by a pixel of a CCD camera is representative of the concentration of the fluorophores on the cell surface imaged by that pixel, the polarization (or anisotropy, *r*) values of two pixels differing in fluorescence intensity will be different if the fluorophores are within energy-transfer distances of each other. If GPI-anchored proteins are arranged in subpixel-sized domains or 'rafts' at the cell surface (**b**), the local fluorophore density would be proportional to the number of such domains in a given pixel, and the distance between fluorophores would be independent of the fluorophore density. If these domains are submicron-sized, the anisotropy values of two pixels differing in fluorescence intensity will be independent of fluorophore intensity.

surface (Fig. 1b). To distinguish between these possibilities, we determined whether there was any change in anisotropy by increasing the average distance between fluorophores by diluting the fluorescent label with folic acid or by photobleaching a given field for different times, and found that the anisotropy increased with increasing photobleaching times (Fig. 3a). Anisotropy values for pixels having the same fluorescence intensities are dependent on the extent of bleaching, consistent with the proposal that the GPI-linked species are in sub-pixel-sized domains (Fig. 1b). Similar increases in anisotropy have been obtained with serial dilution of the label with folic acid: the mean anisotropy was 0.244 ± 0.0071 , 0.267 ± 0.0059 , 0.274 ± 0.0074 for the undiluted, 1:1 diluted, and 1:4 diluted label, respectively. This effect is not specific to CHO cells because results were similar for folate-receptor organization in Caco-2 cells (Fig. 3b). These data prove that homotransfer of

energy between PLF-labelled FR-GPI species takes place at all fluorescence intensities, irrespective of their concentration in the membrane.

In contrast, photobleaching of PLF-labelled FR-TM-expressing cells showed that anisotropy values of pixels having the same fluorescence intensity were independent of the extent of photobleaching (Fig. 3c), consistent with a random organization (Fig. 1a). Furthermore, in transient transfections we determined the anisotropy for PLF-labelled FR (GPI and TM forms) at similar fluorescence intensity values, with identical results (Fig. 2, and data not shown). This difference between the organization of GPI-anchored and TM-anchored folate receptor is therefore not dependent on a particular clone of folate-receptor-expressing cells. We have detected the presence of other GPI-anchored proteins in the same domain, confirming that these experiments do not probe homotypic interactions between like GPI-anchored proteins (R.V. and S.M., unpublished observations). Together, our results indicate that the GPI anchor is necessary for organizing these proteins into sub-pixel-sized domains at the surface of living cells.

Assuming that all GPI-linked folate receptors are organized in sub-pixel domains, the maximum distance between individual fluorophores would have to be ≤ 10 nm (Forster's length for fluorescein bound to a protein is ~ 5.3 nm; ref. 21) for there to be significant energy transfer between individual molecules. As each pixel captures a fluorescence signal from an ensemble of domains and the extent of depolarization is independent of fluorescence intensity over a wide range (200-fold) of values, the size of an individual domain is $1/\sqrt{(200)}$ the size of a pixel ($1 \mu\text{m}$). Therefore, these domains are likely to be smaller than 70 nm, and contain fewer than 50 molecules of GPI-anchored proteins.

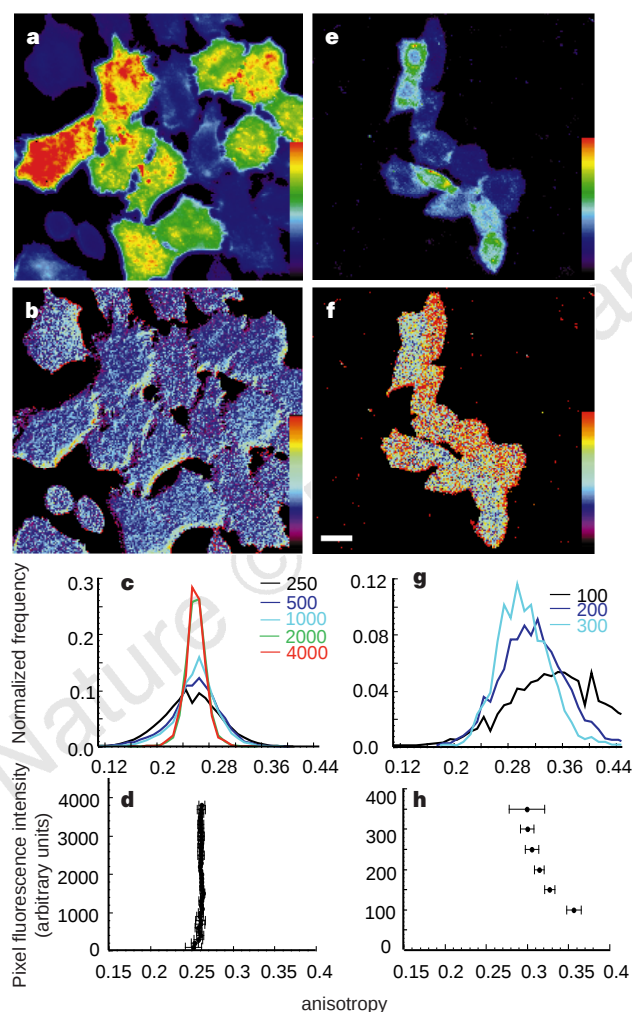


Figure 2 Distribution of fluorescence intensity and anisotropy values at the surface of PLF-labelled FR-GPI- and FR-TM-expressing cells. Cells expressing FR-GPI (a-d) and FR-TM (e-h) were labelled with PLF at the cell surface and imaged as described in Methods. Total intensity (a, e) and anisotropy (b, f) images from a single field of cells are shown as pseudo-coloured 8-bit images. Anisotropy distributions of fluorescence intensity selected pixels from FR-GPI (c: 250 ± 50 , black; 500 ± 50 , purple; $1,000 \pm 50$, blue; $2,000 \pm 50$, green; $4,000 \pm 50$, red), or FR-TM (g: 100 ± 50 , black; 200 ± 50 , purple; 300 ± 50 , blue) were obtained from 75 cells in each case, and the frequency was normalized to the number of pixels in each intensity slice. Mean values of anisotropy and standard deviations at different intensities (± 50 units) were obtained from a minimum of four fields, and plotted against the corresponding pixel intensity values as described in Methods (d, h). Scale bar, $20 \mu\text{m}$.

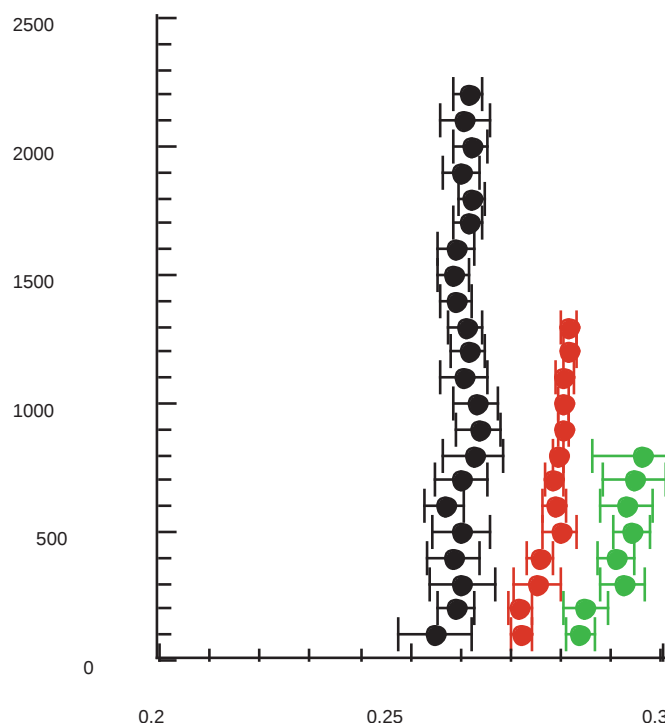


Figure 3 Anisotropy distributions at the surface of PLF-labelled cells obtained after photobleaching. CHO cells expressing FR-GPI (a) or FR-TM (c), or Caco-2 cells (b) were labelled with PLF and subsequently photobleached for 0 (black), 30 (red) or 60 (green) successively for the same field of cells. Mean anisotropy values and standard deviations at different intensities were determined from at least two fields and plotted against the corresponding pixel intensity values, as described in Methods.

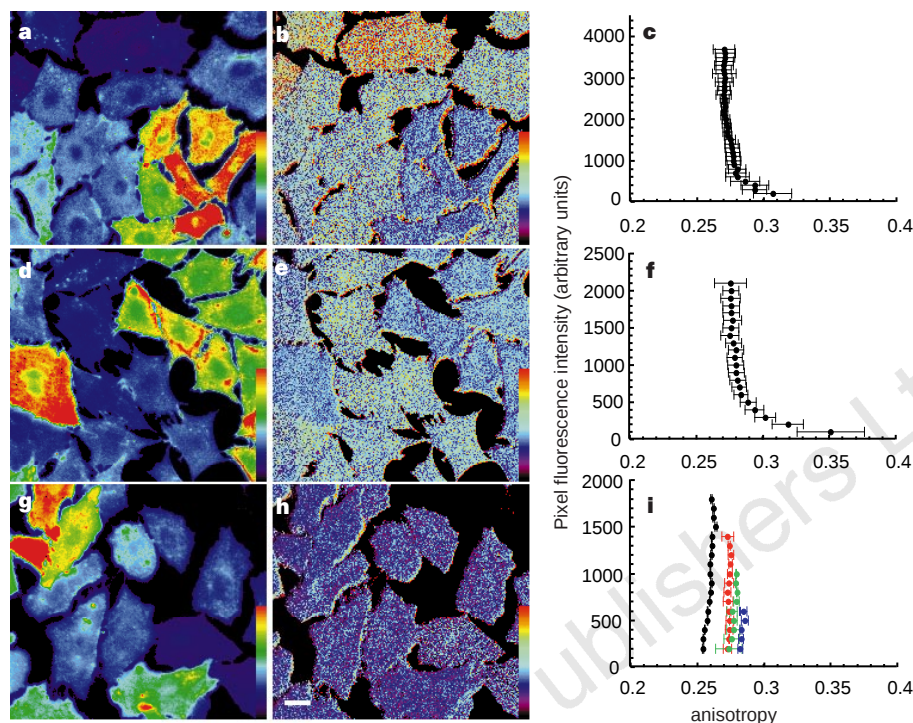


Figure 4 Effect of cholesterol depletion on the organization of GPI-anchored proteins at the surface of cells. FR-GPI-expressing cells were labelled with PLF at the cell surface and depleted of cholesterol by treatment with saponin at low temperatures (**a–c**), or first depleted of cholesterol by treatment with compactin (**d–f**) and then replenished with exogenous cholesterol (**g–i**) before being labelled with PLF. Total intensity (**a, d, g**) and anisotropy (**b, e, h**) images for a single field of cells are shown as pseudo-coloured images. Mean anisotropy values and

standard deviations obtained at different fluorescence intensities were determined from at least 8 fields from saponin (**c**), 7 fields from compactin-treated (**f**), and 3 fields for the cholesterol-replenished (**i**) cells, and plotted against the corresponding pixel fluorescence values as described in Methods. The cholesterol-replenished cells (**i**) were subsequently photobleached for 0 (black), 30 (red), 60 (green) and 90 (purple) successively for the same field of cells. Scale bar, 20 μm .

Depletion of cellular cholesterol impairs the ability of GPI-anchored proteins to associate with the detergent-insoluble membrane fraction^{22,23}, to undergo conversion to an infective scrapie form in endosomes²⁴, to take up folates efficiently²⁵, to be retained in endosomes²⁶, or to transduce signals in lymphocytes²⁷. As cholesterol in artificial membranes is crucial for the formation of domains (reviewed in refs 3, 6, 7), these effects may be due to the disintegration of GPI-anchored protein domains as a result of cholesterol depletion. We therefore depleted the cell membrane of cholesterol by treatment with saponin at low temperatures²² (Fig. 4a–c), or by incubation with methyl- β -cyclodextrin²⁸ (data not shown) or compactin²⁵ (Fig. 4d–f). In all cases, there was an increase in the average anisotropy for all fluorescence intensities relative to undepleted cells (Figs 2c, 3a). This increase in anisotropy is consistent with a decrease in FRET efficiency between GPI-anchored species and an increase in the average distance between molecules. It is unlikely to be due to a change in rotational diffusion rates of the protein because anisotropy values of individual pixels in cholesterol-depleted cells are inversely dependent on fluorescence intensity, consistent with a disruption of domain organization of FR-GPI. In contrast to photobleaching profile of FR-TM (Fig. 3c), we saw a shift in anisotropy values after initial photobleaching, indicating that not all the domains are disrupted (data not shown). Addition of cholesterol to depleted cells restored the organization of GPI-anchored protein in domains; anisotropy of replenished cells was reduced and became independent of fluorescence intensity, and the photobleaching profiles (Fig. 4i) were like those of untreated cells (Fig. 3a). Cholesterol depletion had no effect on the anisotropy values of PLF-labelled FR-TM-expressing cells (data not shown). These results show that cholesterol organizes domains of GPI-anchored proteins in living cell membranes.

The close-packing of GPI-anchored proteins in cholesterol-dependent domains explains why the function and/or sorting of these proteins is dictated by the presence of a GPI-anchor and cholesterol^{24–26}. Significant local concentrations of the scrapie prion protein or the folate receptor in the milieu of the membrane are probably necessary for efficient conversion to the infective form²⁴ or folate uptake²⁹, respectively. The inhibition caused by cholesterol depletion of signalling in lymphocytes that is mediated by GPI-anchored proteins²⁷ is consistent with these proteins associating with preassembled signalling molecules in cholesterol-dependent domains¹⁴.

Our results, together with those in an accompanying paper³⁰, show that lipid-anchored molecules such as GPI-anchored proteins are arranged in submicron-sized domains at the surface of cells, confirming the raft hypothesis. The FRET technique used in conjunction with appropriate protein chimaeras tagged with green fluorescent protein, should reveal the molecular arrangement of many proteins, including cytoplasmic proteins in living cells. □

Methods

Cells and cell culture. CHO cells expressing GPI-anchored folate receptor (FR-GPI) were derived as described⁹. Cells expressing transmembrane anchored folate receptor (FR-TM) were derived from a CHO line by transfection with pMFR-FL1, a plasmid constructed from the eukaryotic expression vector pMEP4 (Invitrogen). pMFR-FL1 consists of the *Hind*III/*Pst*I folate-binding fragment of the cDNA of the human folate receptor²⁹ fused to the transmembrane sequence of the mouse IgG Fc receptor (Fc) receptor and cytoplasmic tail of the low-density lipoprotein (LDL) receptor obtained by PCR-based cloning techniques from the plasmid pFcLR5-50 (ref. 31). The coding region of the chimaeric receptor in pMFR-FL1 was

sequenced and found to contain an intact folate-binding domain with a single conservative mutation of Val to Ala at position 78 and unchanged Fc and LDL sequences. The transfected FR-TM-expressing cells bound PLF with the same affinity as the FR-GPI-expressing cells. All FR-expressing cells were labelled at 0 °C with 10 nM PLF when they were 70–80% confluent, as described⁹. Transient transfections in CHO cells were carried out using lipofectamine (Life Technologies) as a DNA carrier, essentially according to the manufacturers' recommendations. Frequently, more than 70% of the cells in a given field were found to express the folate-binding protein and this was confirmed by the specific binding of an anti-folate receptor antibody, Mov-19 (ref. 9).

Cholesterol depletion. Cells were depleted of cholesterol by three independent means. PLF-labelled FR-GPI-expressing cells were treated with 0.2% of saponin in medium 1 on ice for 30 min²² before fluorescence microscopy. Alternatively, cells were incubated with 10 mM methyl- β -cyclodextrin at 37 °C for 1 hour²⁸, or treated with compactin (10 μ M) for 4 days in the presence of lipoprotein-depleted serum and 200 μ M mevalonate, and then labelled with PLF at 0 °C as described above. Compactin treatment or cyclodextrin treatment resulted in a comparable reduction of 40 or 32% in cholesterol levels, respectively, in CHO cells (data not shown), similar to that obtained with MA104 cells²⁵. Compactin-treated cells were replenished with cholesterol by incubating with dialysed serum containing lipoproteins in the presence of compactin 12 h before labelling with PLF.

Fluorescence microscopy and anisotropy measurements. All fluorescence experiments were done on a Nikon TMD fluorescence microscope equipped with a cooled back-illuminated, 16-bit CCD camera. PLF images were collected using fluorescein optics and were acquired with Metamorph software (Universal Imaging). For anisotropy measurements, fluorescence images were collected using polarizers fitted in a filter slider placed in front of the camera, aligned parallel and perpendicular to the polarizer kept in the excitation path (the polarizers were aligned to better than 98.75%, as assayed by reflection measurements off the coverslip). Parallel and perpendicular images were collected successively using the two differently aligned polarizers. Images were collected with 1-s exposure times and the two images were collected within ~4 s of each other. Alignment of the excitation polarizer at 90° did not change the overall result, showing that the fluorophores are not oriented in the membrane. The amount of photobleaching between images was estimated to be <1% and had no effect on the data. Fluorescence imaging was carried out with a 20 $\times$, NA 0.75 objective at a pixel size of 1 μ m or 0.5 μ m. The pixel size is larger than the measured Airy disc for this objective at 510 nm. Identical data were obtained at both these magnifications. Background images were collected from 'blank' dishes containing the same medium and under the same illumination condition as the experimental dishes, and subtracted from the images before analysis. The magnitude of autofluorescence for a given cell is less than 1% for low-expressing cells (fluorescence intensity values of ~200), similar to the values obtained for cells treated with PIPLC (phosphatidylinositol-specific phospholipase C) before labelling.

Total intensity and anisotropy values were calculated using the following formulae: total intensity = $I_{||} + 2I_{\perp}$; anisotropy = $I_{||} - I_{\perp}/I_{||} + 2I_{\perp}$, where I is the intensity and subscript symbols indicate parallel and perpendicular.

Total intensity and anisotropy images were constructed in Metamorph using 16-bit arithmetic. Total intensity images were imported in Adobe Photoshop 4.0, rescaled to fill the entire 8-bit range, and then pseudo-coloured using a look-up-table (LUT) custom-designed to visualize the entire range of intensities. Anisotropy images were imported into Adobe Photoshop and converted to 8-bit images such that the lowest pixel value (0) corresponds to anisotropy value 0.2 and the highest (255) corresponds to value 0.4, and pseudocoloured. The LUT bars in each case were generated from an 8-bit grey scale image and the colours adjusted to give maximum contrast to the range of total intensities and anisotropy values. All images for each type of measurement (intensity or anisotropy) were pseudo coloured using the same pseudocolour LUT.

Data analysis. C programs were written to analyse the data from a number of fields. Anisotropy distributions for selected intensities were generated by calculating anisotropy values for pixels lying within an interval of 50 or 100 of the selected intensity. Each intensity interval contained at least 200 data points.

The weighted mean, μ , from a minimum of four fields was obtained as follows:

$$\mu = \frac{\sum \left(\frac{n_i \mu_i}{\sigma_i^2} \right)}{\sum \left(\frac{n_i}{\sigma_i^2} \right)}$$

where n_i is the number of data points, μ_i the mean, and σ_i the standard deviation of each distribution in each intensity-selected slice from the i th field.

The deviation in the mean, σ , was calculated as $\sigma^2 = \Sigma(\mu - \mu_i)^2/N$, where N is the total number of fields. Error bars in all the figures represent this deviation.

Received 11 May; accepted 17 June 1998.

- Jacobson, K., Sheets, E. D. & Simson, R. Revisiting the fluid mosaic model of membranes. *Science* **268**, 1441–1442 (1995).
- Simons, K. & Ikonen, E. Functional rafts in membranes. *Nature* **387**, 569–570 (1997).
- Harder, T. & Simons, K. Caveolae, DIGs, and the dynamics of sphingolipid-cholesterol microdomains. *Curr. Opin. Cell Biol.* **9**, 534–542 (1997).
- Weimbs, T., Hui-Low, S., Chapin, S. J. & Mostov, K. E. Apical targeting in polarized cells: there's more afloat than rafts. *Trends Cell Biol.* **7**, 393–399 (1997).
- Weber, G. Dependence of polarization of the fluorescence on the concentration. *Trans. Faraday Soc.* **50**, 552–555 (1954).
- Edidin, M. Lipid microdomains in cell surface membranes. *Curr. Opin. Struct. Biol.* **7**, 528–532 (1997).
- Brown, R. E. Sphingolipid organization in biomembranes: what physical studies of model membranes reveal. *J. Cell Sci.* **111**, 1–9 (1998).
- Kurzchalia, T. V., Hartmann, E. & Dupree, P. Guilt by insolubility—does a protein's detergent insolubility reflect a caveolar location? *Trends Cell Biol.* **5**, 187–189 (1995).
- Mayor, S. & Maxfield, F. R. Insolubility and redistribution of GPI-anchored proteins at the cell surface after detergent treatment. *Mol. Biol. Cell* **6**, 929–944 (1995).
- McConville, M. J. & Ferguson, M. A. The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. *Biochem. J.* **294**, 305–324 (1993).
- Mayor, S., Rothberg, K. G. & Maxfield, F. R. Sequestration of GPI-anchored proteins in caveolae triggered by cross-linking. *Science* **264**, 1948–1951 (1994).
- Parton, R. G., Joggerst, B. & Simons, K. Regulated internalization of caveolae. *J. Cell Biol.* **127**, 1199–1215 (1994).
- Fujimoto, T. GPI-anchored proteins, glycosphingolipids, and sphingomyelin are sequestered to caveolae only after crosslinking. *J. Histochem. Cytochem.* **44**, 929–941 (1996).
- Brown, D. The tyrosine kinase connection: how GPI-anchored proteins activate T cells. *Curr. Opin. Immunol.* **5**, 349–354 (1993).
- Griffiths, G. *Fine Structure Immunocytochemistry* 1–459 (Springer, Heidelberg, 1993).
- Simson, R. et al. Structural mosaicism on the submicron scale in the plasma membrane. *Biophys. J.* **74**, 297–308 (1998).
- Sheets, E. D., Lee, G. M., Simson, R. & Jacobson, K. Transient confinement of a glycosylphosphatidylinositol-anchored protein in the plasma membrane. *Biochemistry* **36**, 12449–12458 (1997).
- Kenworthy, A. K. & Eddin, M. Searching for 'lipid rafts' in cell membranes using fluorescence resonance energy transfer (FRET) microscopy. *Biophys. J.* **74**, A8 (1998).
- Matko, J. & Edidin, M. Energy transfer methods for detecting molecular clusters on cell surfaces. *Meth. Enzymol.* **278**, 444–462 (1997).
- Hannan, L. A., Lisanti, M. P., Rodriguez-Boulant, E. & Edidin, M. Correctly sorted molecules of a GPI-anchored protein are clustered and immobile when they arrive at the apical surface of MDCK cells. *J. Cell Biol.* **120**, 353–358 (1993).
- Runnels, L. W. & Scarlata, S. F. Theory and application of fluorescence homotransfer to melittin oligomerization. *Biophys. J.* **69**, 1569–1583 (1995).
- Cerneus, D. P., Ueffing, E., Postuma, G., Strous, G. J. & van der Ende, A. Detergent insolubility of alkaline phosphatase during biosynthetic transport and endocytosis. Role of cholesterol. *J. Biol. Chem.* **268**, 3150–3155 (1993).
- Hanada, K., Nishijima, M., Akamatsu, Y. & Pagano, R. E. Both sphingolipids and cholesterol participate in the detergent insolubility of alkaline phosphatase, a glycosylphosphatidylinositol-anchored protein, in mammalian membranes. *J. Biol. Chem.* **270**, 6254–6260 (1995).
- Taraboulos, A. et al. Cholesterol depletion and modification of COOH-terminal targeting sequence of the prion protein inhibit formation of the scrapie isoform. *J. Cell Biology* **129**, 121–132 (1995).
- Chang, W.-J., Rothberg, K. G., Kamen, B. A. & Anderson, R. G. W. Lowering cholesterol content of MA104 cells inhibits receptor-mediated transport of folate. *J. Cell Biol.* **118**, 63–69 (1992).
- Mayor, S., Sabharanjak, S. & Maxfield, F. R. Cholesterol-dependent retention of GPI-anchored proteins in endosomes. *EMBO J.* (in press).
- Stulnig, T. M. et al. Signal transduction via glycosyl phosphatidylinositol-anchored proteins in T cells is inhibited by lowering cellular cholesterol. *J. Biol. Chem.* **272**, 19242–19247 (1997).
- Yancey, P. G. et al. Cellular cholesterol efflux mediated by cyclodextrins. Demonstration of kinetic pools and mechanism of efflux. *J. Biol. Chem.* **271**, 16026–16034 (1996).
- Ritter, T. E., Fajardo, O., Matsue, H., Anderson, R. G. & Lacey, S. W. Folate receptor targeted to clathrin-coated pits cannot regulate vitamin uptake. *Proc. Natl Acad. Sci. USA* **92**, 3824–3828 (1995).
- Friedrichson, T. & Kurzchalia, T. V. Microdomains of GPI-anchored proteins in living cells revealed by crosslinking. *Nature* **394**, 802–805 (1998).
- Matter, K., Hunziker, W. & Mellman, I. Basolateral sorting of LDL receptor in MDCK cells: the cytoplasmic domain contains two tyrosine-dependent targeting determinants. *Cell* **71**, 741–753 (1992).

Acknowledgements. We thank M. K. Mathew and G. Krishnamoorthy for their insight, F. Barrantes, A. K. Menon and J. Udgankar for critical reading of the text, S. Chitkala and N. S. Kiran for making the FR-TM construct, and R. Priya for help with cholesterol analysis. S.M. thanks the Human Frontier Science Program and the National Centre for Biological Sciences for support, and F. F. Bosphorus for inspiration.

Correspondence and requests for materials should be addressed to S.M. (e-mail: mayor@ncbs.tifrbng.res.in).

Microdomains of GPI-anchored proteins in living cells revealed by crosslinking

Tim Friedrichson & Teymuraz V. Kurzchalia

Department of Cell Biology, Max Delbrück Centre for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin-Buch, Germany

There is some discussion as to whether glycosyl-phosphatidylinositol (GPI)-anchored proteins occur in microdomains in the cell membrane^{1,2}. These putative microdomains have been implicated in processes such as sorting in polarized cells³⁻⁵ and signal transduction⁶⁻⁸. Complexes enriched in GPI-anchored proteins, cholesterol and glycosphingolipids have been isolated from cell membranes by using non-ionic detergents: these complexes were thought to represent a clustered arrangement of GPI-anchored proteins^{9,10}. However, results obtained when clustering of GPI-anchored proteins induced by antibodies or by detergents was prevented support the idea of a dispersed surface distribution of GPI-anchored proteins at steady state¹¹⁻¹³. Here we use chemical crosslinking to show that membrane microdomains of a GPI-anchored protein exist at the surface in living cells. This clustering is specific for the GPI-anchored form, as two transmembrane forms bearing the same ectodomain do not form oligomers. Depletion of membrane cholesterol causes the clustering of GPI-anchored proteins to break up, whereas treatment of cells with detergent substantially increases the size of the complexes. We find that in living cells these GPI-anchored proteins reside in

microdomains consisting of at least 15 molecules, which are much smaller than those seen after detergent extraction.

To study the association of GPI-anchored proteins on the cell surface of intact cells, we used growth hormone (GH) with the GPI-anchoring signal from decay-accelerating factor (DAF) attached to its carboxy terminus (GH-DAF) (Fig. 1a). This molecule is an established model for a GPI-anchored protein¹⁴. When MDCK cells that permanently express GH-DAF (MDCK GH-DAF) were chemically crosslinked at 4 °C with the membrane-impermeable agent bis(sulphosuccinimidyl)suberate (BS3), a prominent band corresponding to a relative molecular mass of 46K (dimer), and a smear from ~60K to ~300K were detected, suggesting the existence of microdomains. The crosslinking efficiency for 0.5 mM BS3 was 71%. When crosslinked at 37 °C, the pattern was similar (crosslinking efficiency was 86%), indicating that microdomains also exist at physiological temperatures. Increasing concentrations of BS3 (4 °C) or increasing the incubation time with crosslinker (37 °C) shifted the pattern of crosslinked products to high-molecular-mass forms (Fig. 1b). To analyse the oligomers formed, we used the cleavable BS3 analogue 3,3'-dithiobis-(sulphosuccinimidylpropionate) (DTSSP) and two-dimensional (2D) electrophoresis. Electrophoresis in the first dimension gave a crosslinking pattern similar to that obtained with BS3, whereas the second dimension under reducing conditions revealed that the crosslinked oligomers consisted primarily of GH-DAF (arrow in Fig. 1c); some minor interaction partners were also detected (arrowheads in Fig. 1c). Because the length of the spacer arm of BS3 is only 1.14 nm, the detection of crosslinked GH-DAF oligomers indicates that these molecules are close together on the cell surface at steady state. We estimate that the domains consist of at least 15 GH-DAF molecules, because the largest crosslinked species corresponds to a multimer of 15 GH-DAF species. However, the number of GH-DAF molecules in a cluster cannot be defined precisely with this approach: the number

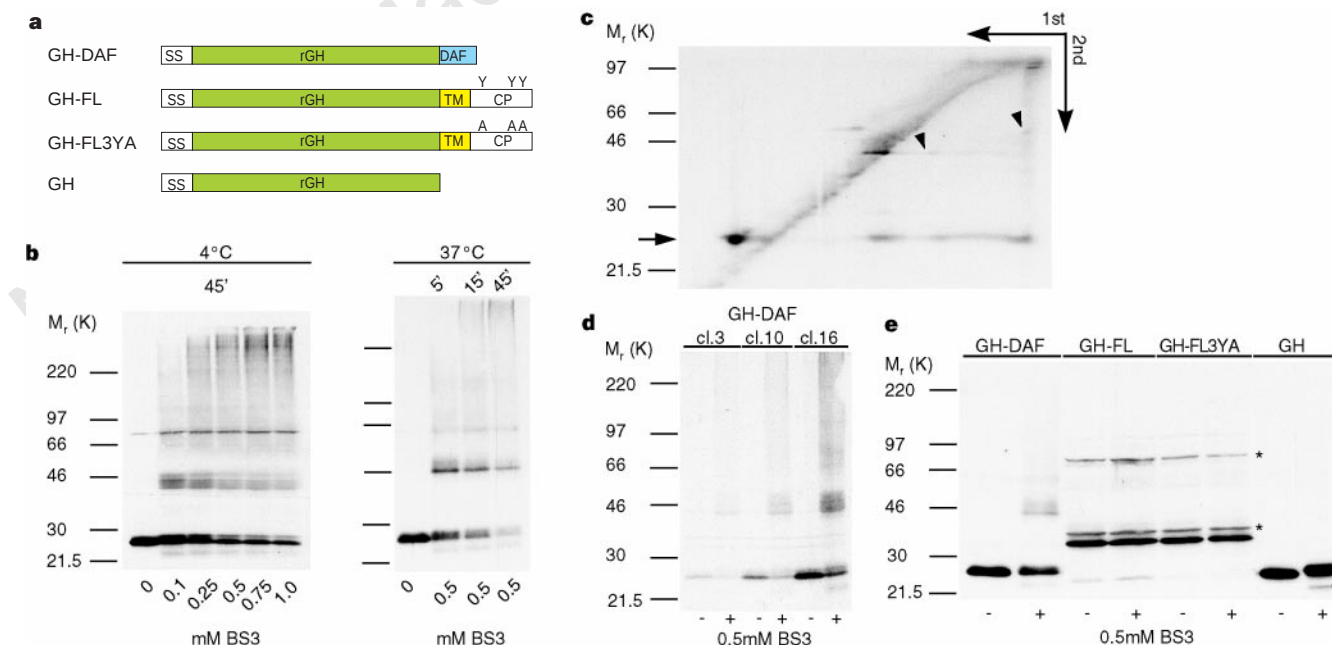


Figure 1 Detection of oligomers upon crosslinking of GPI-anchored, transmembrane and secretory forms of GH. **a**, Constructs used for generation of permanently expressing MDCK cells. The GPI-anchoring signal of DAF was fused C-terminally to GH. Transmembrane forms were generated by fusing the FcR11-B2 receptor transmembrane domain and the LDL-receptor cytoplasmic domain to GH. Mutation of the tyrosine residues present in the C-tail to alanine results in apical expression in polarized MDCK cells. **b**, Appearance of GH-DAF oligomers induced by chemical crosslinking. MDCK GH-DAF cells were incubated with increasing concentrations of BS3 at 4 °C or for different times with 0.5 mM BS3 at 37 °C. **c**, Oligomers formed on crosslinking consist primarily of

GH-DAF. MDCK GH-DAF cells were labelled with ³⁵S-methionine, and after crosslinking were subjected to 2D PAGE. GH-DAF is the principal protein found in the oligomers (arrow). Possible interaction partners are indicated by arrowheads. **d**, Crosslinking pattern is independent of the level of GH-DAF expression. Low- (clone 3), medium- (clone 10) or high- (clone 16) expressing clones were used for crosslinking. **e**, Clustering is a specific property of the GPI-anchored form of GH. MDCK GH-DAF clone 10, GH-FL or GH-FL3YA cells expressing comparable amounts of protein, and MDCK GH cells were crosslinked with BS3. Note that oligomers are detected only for GH-DAF. Asterisk, nonspecific bands obtained upon preparation of TX-114-soluble material (see Methods).

is likely to be an underestimate owing to non-saturated cross-linking.

If GH-DAF molecules are crosslinked just because of their relatively high density in cell membranes, there should be a significant difference in crosslinking efficiency in high- and low-expressing clones; but if GH-DAF is confined to microdomains, there should be a high local concentration of molecules and the extent of crosslinking efficiencies should be independent of the amount of expression. We therefore isolated MDCK clones expressing different amounts of GH-DAF (clones 10 and 16 express 5- and 20-fold more than clone 3, respectively) and used them for cross-linking (Fig. 1d). Quantification of the bands on the blot revealed that the crosslinking efficiency was comparable in all clones (86, 81 and 78% for clones 3, 10 and 16, respectively), irrespective of the amount of GH-DAF expressed (Fig. 1d). To compare crosslinking patterns, we used the ratio of high-molecular-mass (top third of crosslinked species) to total crosslinked species. These ratios were 30, 23 and 25% for clones 3, 10 and 16, respectively. For all three clones, crosslinking occurred even when the reaction time was very short (2 or 5 min; data not shown), confirming that the GPI-anchored protein was present in domains with a high local concentration.

To investigate whether clustering is specific for GH-DAF, cell lines bearing a transmembrane form of growth hormone were generated: MDCK GH-FL (growth hormone fused to the transmembrane domain of the FcR2-B2 receptor and the cytosolic domain of the low-density lipoprotein (LDL) receptor; Fig. 1a) and MDCK GH-FL3YA (in which three cytosolic tyrosine residues have been mutated to alanine). Cells expressing comparable amounts of protein (MDCK GH-FL and GH-FL3YA cells express 73 and 76% of the amount of protein expressed in GH-DAF cl. 10 cells, respectively) were crosslinked with BS3. In contrast to GH-DAF, the amount of GH-FL and GH-FL3YA monomers remaining after crosslinking was unchanged and no oligomers were formed (Fig. 1e); the secretory form of GH would not crosslink under a variety of conditions (Fig. 1e, and data not shown). Together, these findings indicate that the propensity of GH-DAF to cluster is due to its GPI anchor and not to its protein moiety. Note that GH-DAF and GH-FL3YA are both expressed apically in polarized MDCK cells (data not shown), indicating that microclustering and apical sorting in epithelial cells may not be connected.

Is the clustered organization at the plasma membrane a general property of GPI-anchored proteins? When other cells (CHO) that permanently express another GPI-anchored protein (the folate receptor, CHO FR-GPI) were crosslinked with BS3, the pattern of

crosslinked products was similar to that obtained with MDCK GH-DAF (dimers, trimers and so on) (Fig. 2b).

Although the existence of GPI-anchored protein clusters in the plasma membrane of living cells is unproven^{2,15}, cholesterol depletion leads to a loss of antibody-induced clustering of GPI-anchored proteins¹⁶, an increase in their detergent solubility^{17,18}, inhibition of folate uptake^{19,20} and of signal transduction by GPI-anchored proteins²¹. To see whether cholesterol depletion affected microclustering of GH-DAF and FR-GPI, we treated MDCK GH-DAF cells with methyl- β -cyclodextrin²² and found that the cellular cholesterol content was reduced to 37% of that of control cells, causing the crosslinking of GH-DAF to be inhibited (Fig. 2a) and the efficiency to be reduced to 14%; also, the clustering of FR-GPI in CHO FR-GPI cells was almost abolished after extraction of cholesterol (Fig. 2b). Results were the same, irrespective of the cholesterol-depletion protocol. Replenishing cholesterol-depleted MDCK GH-DAF cells with cholesterol increased their cholesterol content to 78% and partly restored crosslinking (crosslinking efficiency increased to 25%; Fig. 2a). These results indicate that cholesterol-dependent clustering of GH-DAF is partially reversible and in principle could be modulated in the cell, and that the dependence of the integrity of the GPI-anchored protein cluster on cholesterol seems to be a general phenomenon not confined to particular cell types.

The identification of the lectin-like membrane protein VIP36 in detergent-insoluble complexes²³ and of *N*-glycans as possible sorting signals for secreted proteins in MDCK cells²⁴ has raised the question of whether carbohydrate moieties are involved in transport and/or sorting of GPI-anchored protein. As the GPI anchor consists of both lipid and carbohydrate, we investigated whether carbohydrates are needed to maintain oligomers by crosslinking MDCK GH-DAF and CHO FR-GPI cells in the presence of different monosaccharides (glucose, mannose, inositol and *N*-acetylglucosamine), assuming that excess structural carbohydrates from the GPI anchor might interfere with crosslinking. *N*-acetylglucosamine was used instead of glucosamine, which contains a reactive amino group that would prevent crosslinking. In contrast to the results from cholesterol-depletion experiments, we detected no changes in the pattern of crosslinked products (data not shown).

When the cell membrane is extracted with non-ionic detergents, a fraction enriched in cholesterol, glycosphingolipids and GPI-

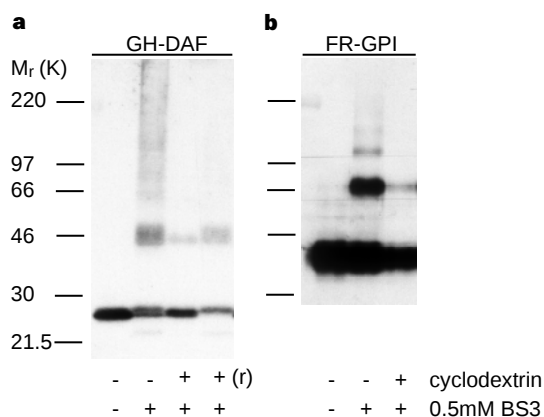


Figure 2 Clustering of GPI-anchored proteins depends on cholesterol. **a**, MDCK GH-DAF cells, or **b**, CHO FR-GPI cells were treated with 10 mM methyl- β -cyclodextrin for 1 h at 37°C (where indicated) to deplete the membrane of cholesterol. MDCK GH-DAF cells were partially replenished with cholesterol by incubation with DMEM supplemented with 10% FCS for 6 h at 37°C. After these treatments, cells were crosslinked with BS3.

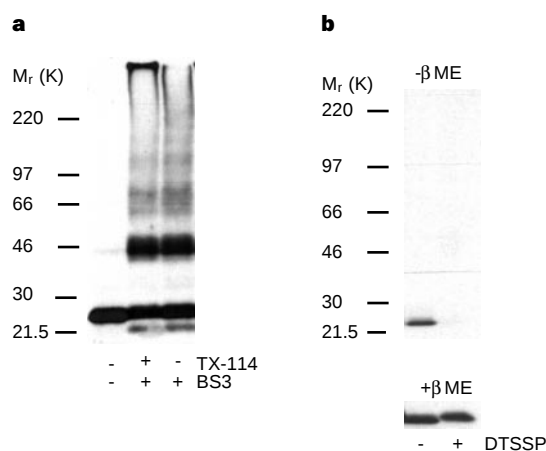


Figure 3 Detergent treatment leads to increased size of crosslinked oligomers. **a**, Mild TX-114 extraction shifts the oligomerization pattern of GH-DAF to high-molecular-weight forms. Cells were treated with 0.5% TX-114 for 5 min on ice before crosslinking with BS3. **b**, Harsh extraction leads to extreme aggregation of GH-DAF. Cells were extracted for 30 min on ice with 1% TX-114 and the soluble and insoluble fractions separated by centrifugation for 15 min at 15,000g. The insoluble fraction was crosslinked with DTSSP and, after phase separation, the detergent phases were resolved under non-reducing (top) or reducing (bottom) conditions.

anchored proteins can be isolated^{1,4,10}, forming a basis for arguments for the existence of rafts¹. Others, however, claim that GPI-anchored proteins aggregate only after addition of detergents or antibodies^{12,13,15,25}. We investigated the connection between our crosslinked oligomers and the clusters found after detergent extraction of cells by crosslinking MDCK GH-DAF cells after extraction with Triton-X-114 detergent using mild or harsh conditions (see Methods; the latter is commonly used for preparing detergent-insoluble complexes). Mild TX-114 extraction leads to a shift to high-molecular-weight crosslinked products (Fig. 3a); this shift is more evident in samples subjected to harsh TX-114 extraction: complexes are so large that only a small fraction (10%) can enter the gel (Fig. 3b, top). Upon cleavage of the crosslinker DTSSP with a reducing agent, 95% of the material is recovered (Fig. 3b, bottom). These results indicate that the crosslinked clusters of GH-DAF are significantly smaller than the clusters induced by Triton X-114 treatment: this could be due to the merging of small microdomains⁴.

According to the hypothesis of Simons and Ikonen¹, glycosphingolipids and cholesterol can form platforms (rafts) with which some proteins, such as GPI-anchored proteins, are closely associated and from which other membrane proteins are excluded; however, the existence of rafts in living cells, particularly of any containing GPI-anchored proteins, has not been demonstrated. Clustering of GPI-anchored proteins may have been overestimated or small clusters may have escaped detection because of limitations of the microscopical methods used¹². The results from a fluorescence resonance energy-transfer technique²⁶ together with those presented here, to our knowledge provide the first evidence for the existence of rafts *in vivo*: we conclude the GPI-anchored proteins reside in membrane microdomains which need cholesterol to be maintained, and that these proteins become heavily aggregated after detergent extraction. We shall not know the exact size or the protein and lipid composition of the rafts until new methods can be devised to isolate rafts without using detergents. This isolation will be difficult because rafts are probably dynamic structures that are constantly exchanging their components²⁷. For instance, the discovery of caveolin-1, a protein of the caveolar coat, in detergent extracts enriched in glycosphingolipids, cholesterol and GPI-anchored proteins has led to the proposed existence of rafts containing caveolin-1 and GPI-anchored proteins¹⁰. However, there is accumulating evidence that there is no enrichment of GPI-anchored proteins in caveolae^{25,28,29}; the same is true for trimeric G proteins, which are also found in a buoyant raft fraction but are neither functionally nor morphologically connected with caveolin³⁰. The rules governing the interplay between the inclusion and exclusion of proteins into rafts remain to be defined. □

Methods

Cloning. PCR cloning techniques were used to replace the stop codon of rat growth hormone with a *Bam*HI site creating pSP64GHΔC. The same site was added upstream of the GPI-anchoring signal from DAF and cloned into pSP64GHΔC to yield pSP64GH-DAF. Plasmid pCBFL5-50 encodes the FcR2-B2 receptor ecto- and transmembrane domains fused to the LDL-receptor cytosolic domain³¹. In pCBFL5-503YA, the three cytosolic tyrosine residues are mutated to alanine. PCR was used to add *Bam*HI sites upstream of the transmembrane sequences, and fragments were cloned in pSP64GHΔC to create pSP64GH-FL and pSP64GH-FL3YA, respectively. All constructs were inserted into the mammalian expression vector pRc/CMV (Invitrogen).

Cells, cell culture and transfection. MDCK cells were maintained in DMEM supplemented with 10% FCS and antibiotics. CHO FR-GPI cells were maintained in folate-free Hams F-12 medium containing 5% FCS, antibiotics and 100 μg ml⁻¹ hygromycin. For permanent transfection of MDCK cells, the calcium phosphate precipitation method was used.

Crosslinking, electrophoresis and western blotting. Cells were washed twice with ice-cold PBS and chilled on ice. BS3 or DTSSP were diluted from 25 mM stocks to 0.5 mM with PBS unless indicated otherwise. Except where noted, crosslinking was quenched after 45 min by addition of 50 mM glycine for 15 min. All incubations were at 4 °C. MDCK GH-DAF cells (clone 16, except where noted)

were washed with PBS and lysed for 20 min at 4 °C and 10 min at 37 °C in TX-114 lysis buffer (150 mM NaCl; 10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 1% Triton X-114, and protease inhibitors). Lysates were collected, chilled on ice and cleared by 15 min centrifugation at 15,000g. Cleared lysates were subjected to temperature-induced phase separation for 5 min at 37 °C. Aqueous and detergent phases were separated by centrifugation for 3 min at 13,000 r.p.m. at room temperature. To the detergent phases, 0.9 ml Triton X-114 wash buffer (0.06% TX-114, otherwise as TX-114 lysis buffer) was added and vortexed before centrifugation for 15 min at 15,000g (at 4 °C) and another round of phase separation. Detergent phases were precipitated with cold acetone (-20 °C) and boiled for 5 min at 95 °C in Laemmli sample buffer. MDCK GH-FL and GH-FL3YA cells were crosslinked and lysed in TX-114 lysis buffer for 30 min on ice. Cells were scraped and passed three times through a 26-gauge needle to shear the DNA. After 15 min centrifugation at 15,000g at 4 °C, soluble fractions were precipitated with trichloroacetic acid (TCA). MDCK GH cells were incubated for 2 h with PBS containing 0.9 mM Ca²⁺ and 0.5 mM Mg²⁺ (PBS(+)) and BS3 was added to 0.5 mM directly from a 25 mM stock. Proteins in the supernatant were precipitated with TCA. Pellets were washed with acetone and boiled for 5 min at 95 °C in Laemmli sample buffer. Samples were resolved on 5–15% SDS–PAGE, transferred to nitrocellulose and detected with polyclonal antibodies against GH or folate receptor, HRP-labelled secondary antibodies and ECL. For 2D PAGE, MDCK GH-DAF cells were labelled for 12 h with 80 μCi ³⁵S-methionine (NEN) and crosslinked with DTSSP. The precipitated detergent phase was boiled in non-reducing sample buffer and resolved on 5–15% SDS–PAGE (1D). The lane was cut and boiled in sample buffer containing 5% β-mercaptoethanol for 5 min at 95 °C and then run on a 5–15% gel (2D) and autoradiographed.

Cholesterol depletion and detergent treatments. Cells were depleted of membrane cholesterol with 10 mM methyl-β-cyclodextrin in PBS(+) for 1 h at 37 °C. Replenishing with cholesterol was for 6 h at 37 °C with DMEM containing 10% FCS. Cellular cholesterol was measured with an enzyme test kit (Boehringer Mannheim). Protein was determined using a BCA protein assay kit (Pierce). Detergent extractions were done as follows. After washing cells twice with PBS and chilling on ice, samples subjected to mild TX-114 treatment were extracted for 5 min at RT with 0.5% TX-114 in PBS, washed with PBS twice and crosslinked. Harsh TX-114 treatment was done by lysing PBS-washed cells for 30 min at 4 °C in 1 ml TX-114 lysis buffer; cells were scraped and lysates centrifuged for 15 min at 15,000g. The pellets were treated with 950 μl TX-114 lysis buffer and samples were incubated for 10 min at 37 °C, chilled on ice and crosslinked with BS3 or DTSSP being added directly from 25 mM stocks.

Quantification of western blots. Quantification of western blots was done with a FUJIX BAS-2000 (Fuji) using ³⁵S-labelled anti-rabbit IgGs (Amersham) as secondary reagents, or by scanning of low-exposure images after ECL detection using TINA version 2.08 software (Raytest Isotopenmeßgeräte GmbH). Cross-linking efficiency for blots incubated with ³⁵S-labelled IgG was calculated after background subtraction with the formula: (total p.s.l. – p.s.l. of monomer)/(total p.s.l.), where p.s.l. is a measure of the radiation dose that is proportional to d.p.m. For scanned images, optical density was the unit of measurement.

Received 12 February; accepted 17 June 1998.

1. Simons, K. & Ikonen, E. Functional rafts in cell membranes. *Nature* **387**, 569–572 (1997).
2. Weimbs, T., Low, S. H., Chapin, S. J. & Mostov, K. E. Apical targeting in polarized epithelial cells: there's more afloat than rafts. *Trends Cell Biol.* **7**, 393–399 (1997).
3. Lisanti, M. P., Le Bivic, A., Sargiacomo, M. & Rodriguez Boulan, E. Steady-state distribution and biogenesis of endogenous Madin–Darby canine kidney glycoproteins: evidence for intracellular sorting and polarized cell surface delivery. *J. Cell Biol.* **109**, 2117–2127 (1989).
4. Brown, D. A. & Rose, J. K. Dorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell* **68**, 533–544 (1992).
5. Kurzchalia, T. V. et al. VIP21, a 21-kD membrane protein is an integral component of trans-Golgi-network-derived transport vesicles. *J. Cell Biol.* **118**, 1003–1014 (1992).
6. Stefanova, I., Horejsi, V., Anstegui, I. J., Knapp, W. & Stockinger, H. GPI-anchored cell-surface molecules contribute to protein tyrosine kinases. *Science* **254**, 1016–1019 (1991).
7. Brown, D. The tyrosine kinase connection: how GPI-anchored proteins activate T cells. *Curr. Opin. Immunol.* **5**, 349–354 (1993).
8. Rodgers, W., Crise, B. & Rose, J. K. Signals determining protein tyrosine kinase and glycosyl-phosphatidylinositol-anchored protein targeting to a glycolipid-enriched membrane fraction. *Mol. Cell Biol.* **14**, 5384–5391 (1994).
9. Fiedler, K., Kobayashi, T., Kurzchalia, T. V. & Simons, K. Glycosphingolipid-enriched, detergent-insoluble complexes in protein sorting in epithelial cells. *Biochemistry* **32**, 6365–6373 (1993).
10. Sargiacomo, M., Sudol, M., Tang, Z. & Lisanti, M. P. Signal transducing molecules and glycosyl-phosphatidylinositol-linked proteins form a caveolin-rich insoluble complex in MDCK cells. *J. Cell Biol.* **122**, 789–807 (1993).
11. Hannan, L. A., Lisanti, M. P., Rodriguez Boulan, E. & Edidin, M. Correctly sorted molecules of a GPI-anchored protein are clustered and immobile when they arrive at the apical surface of MDCK cells. *J. Cell Biol.* **120**, 353–358 (1993).

12. Mayor, S., Rothberg, K. G. & Maxfield, F. R. Sequestration of GPI-anchored proteins in caveolae triggered by cross-linking. *Science* **264**, 1948–1951 (1994).
13. Mayor, S. & Maxfield, F. R. Insolubility and redistribution of GPI-anchored proteins at the cell surface after detergent treatment. *Mol. Biol. Cell* **6**, 929–944 (1995).
14. Lisanti, M. P., Caras, I. W., Davitz, M. A. & Rodriguez Boulan, E. A glycosphospholipid membrane anchor acts as an apical targeting signal in polarized epithelial cells. *J. Cell Biol.* **109**, 2145–2156 (1989).
15. Kurzchalia, T., Hartmann, E. & Dupree, P. Guilt by insolubility—Does a protein's detergent insolubility reflect caveolar location? *Trends Cell Biol.* **5**, 187–189 (1995).
16. Rothberg, K. G., Ying, Y. S., Kamen, B. A. & Anderson, R. G. Cholesterol controls the clustering of the glycosphospholipid-anchored membrane receptor for 5-methyltetrahydrofolate. *J. Cell Biol.* **111**, 2931–2938 (1990).
17. Cerneus, D. P., Ueffing, E., Posthuma, G., Strous, G. J. & van der Ende, A. Detergent insolubility of alkaline phosphatase during biosynthetic transport and endocytosis. Role of cholesterol. *J. Biol. Chem.* **268**, 3150–3155 (1993).
18. Hanada, K., Nishijima, M., Akamatsu, Y. & Pagano, R. E. Both sphingolipids and cholesterol participate in the detergent insolubility of alkaline phosphatase, a glycosylphosphatidylinositol-anchored protein, in mammalian membranes. *J. Biol. Chem.* **270**, 6254–6260 (1995).
19. Chang, W. J., Rothberg, K. G., Kamen, B. A. & Anderson, R. G. Lowering the cholesterol content of MA104 cells inhibits receptor-mediated transport of folate. *J. Cell Biol.* **118**, 63–69 (1992).
20. Smart, E. J., Mineo, C. & Anderson, R. G. Clustered folate receptors deliver 5-methyltetrahydrofolate to cytoplasm of MA104 cells. *J. Cell Biol.* **134**, 1169–1177 (1996).
21. Stulnig, T. M. *et al.* Signal transduction via glycosyl phosphatidylinositol-anchored proteins in T cells is inhibited by lowering cellular cholesterol. *J. Biol. Chem.* **272**, 19242–19247 (1997).
22. Klein, U., Gimpl, G. & Fahrenholz, F. Alteration of the myometrial plasma membrane cholesterol content with beta-cyclodextrin modulates the binding affinity of the oxytocin receptor. *Biochemistry* **34**, 13784–13793 (1995).
23. Fiedler, K., Parton, R. G., Kellner, R., Etzold, T. & Simons, K. VIP36, a novel component of glycolipid rafts and exocytic carrier vesicles in epithelial cells. *EMBO J.* **13**, 1729–1740 (1994).
24. Scheiffele, P., Peranen, J. & Simons, K. N-glycans as apical sorting signals in epithelial cells. *Nature* **378**, 96–98 (1995).
25. Fujimoto, T. GPI-anchored proteins, glycosphingolipids, and sphingomyelin are sequestered to caveolae only after crosslinking. *J. Histochem. Cytochem.* **44**, 929–941 (1996).
26. Varma, R. & Mayor, S. GPI-anchored proteins are organized in submicron domains at the cell surface. *Nature* **394**, 798–801 (1998).
27. Harder, T. & Simons, K. Caveolae, DIGs, and the dynamics of sphingolipid-cholesterol microdomains. *Curr. Opin. Cell Biol.* **9**, 534–542 (1997).
28. Schnitzer, J. E., McIntosh, D. P., Dvorak, A. M., Liu, J. & Oh, P. Separation of caveolae from associated microdomains of GPI-anchored proteins. *Science* **269**, 1435–1439 (1995).
29. Wu, M., Fan, J., Gunning, W. & Ratnam, M. Clustering of GPI-anchored folate receptor independent of both cross-linking and association with caveolin. *J. Membr. Biol.* **159**, 137–147 (1997).
30. Huang, C. *et al.* Organization of G proteins and adenyl cyclase at the plasma membrane. *Mol. Biol. Cell* **8**, 2365–2378 (1997).
31. Matter, K., Hunziker, W. & Mellman, I. Basolateral sorting of LDL receptor in MDCK cells: the cytoplasmic domain contains two tyrosine-dependent targeting determinants. *Cell* **71**, 741–753 (1992).

Acknowledgements. We thank G. Lewin, M. S. Bhojani and M. Wiedmann for critically reading the manuscript; W. Hunziker for plasmids pCBFL5-50 and pCBFL5-503YA; and A. Henske for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft and Human Frontier Science Program.

Correspondence and requests for materials should be addressed to T.V.K. (e-mail: kurzchal@orion.rz.mdc-berlin.de).

Structure of a cephalosporin synthase

Karin Vålgård[†], Anke C. Terwisscha van Scheltinga[‡], Matthew D. Lloyd[§], Takane Hara[†], S. Ramaswamy[‡], Anastassis Perrakis^{||}, Andy Thompson^{||}, Hwei-Jen Lee[§], Jack E. Baldwin[§], Christopher J. Schofield[§], Janos Hajdu[†] & Inger Andersson[‡]

[†] Department of Biochemistry, Uppsala University, Box 576, S-751 23 Uppsala, Sweden

[‡] Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, S-751 24 Uppsala, Sweden

[§] Oxford Centre for Molecular Sciences and Dyson Perrins Laboratory, Oxford University, South Parks Road, Oxford OX1 3QY, UK

^{||} EMBL Grenoble, c/o ILL, Avenue des Martyrs, BP 156,

F-38042 Grenoble Cedex 9, France

* These authors contributed equally to this work

Penicillins and cephalosporins are among the most widely used therapeutic agents. These antibiotics are produced from fermentation-derived materials as their chemical synthesis is not commercially viable. Unconventional steps in their biosynthesis are catalysed by Fe(II)-dependent oxidases/oxygenases; isopenicillin N synthase (IPNS)^{1,2} creates in one step the bicyclic nucleus of penicillins, and deacetoxycephalosporin C synthase (DAOCS)

catalyses the expansion of the penicillin nucleus into the nucleus of cephalosporins. Both enzymes use dioxygen-derived ferryl intermediates in catalysis but, in contrast to IPNS, the ferryl form of DAOCS is produced by the oxidative splitting of a co-substrate, 2-oxoglutarate (α -ketoglutarate). This route of controlled ferryl formation and reaction is common to many mononuclear ferrous enzymes³, which participate in a broader range of reactions than their well-characterized counterparts, the haem enzymes. Here we report the first crystal structure of a 2-oxoacid-dependent oxygenase. High-resolution structures for apo-DAOCS, the enzyme complexed with Fe(II), and with Fe(II) and 2-oxoglutarate, were obtained from merohedrally twinned crystals. Using a model based on these structures, we propose a mechanism for ferryl formation.

In 1945, from sea water near a sewage outflow on Sardinia, Giuseppe Brotzu, a bacteriologist and local politician, isolated a new species of *Cephalosporium* (named *C. acremonium*), which secreted material with strong antibacterial activity⁴. It was later established that this fungus produced several different antibiotics. The compound in the mixture arbitrarily named cephalosporin C (ref. 5) became the first known member of a new family of antibiotics that are now called cephalosporins. Chemical and X-ray studies established that cephalosporin C contained a β -lactam ring^{6,7} which, in contrast to penicillins, was resistant to hydrolysis by known penicillinases⁸ (β -lactamases). Bacterial resistance to penicillins was new at the time, and the need to combat resistant strains had just begun to emerge.

The biosynthetic pathways of penicillins and cephalosporins share the first two steps (Fig. 1): the condensation of three L-amino acids to give L- δ -(α -aminoadipoyl)-L-cysteinyl-D-valine (ACV), and the four-electron oxidation of this tripeptide by molecular oxygen catalysed by IPNS to give the first formed penicillin, isopenicillin N (refs 9, 10). The committed step in cephalosporin biosynthesis is the expansion of the five-membered thiazolidine ring of the penicillin nucleus into the six-membered

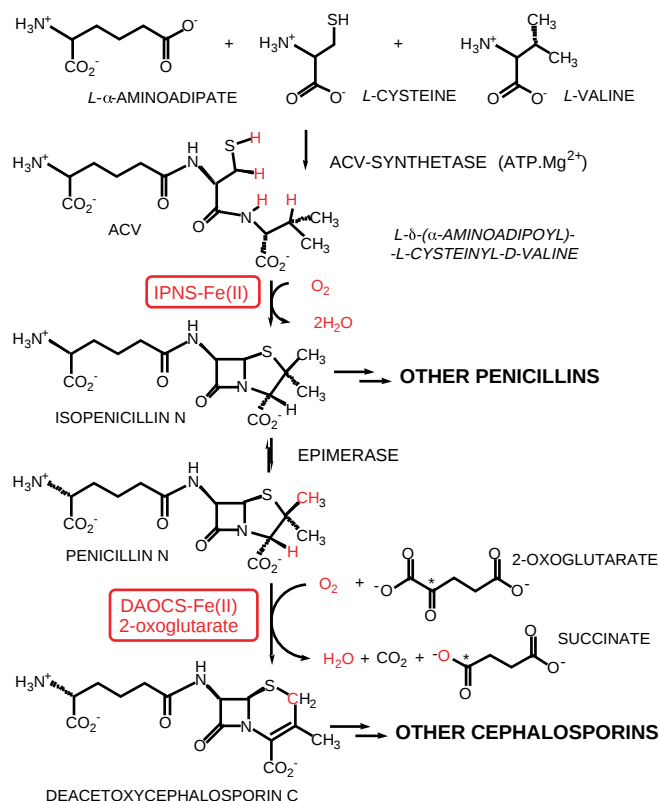


Figure 1 Scheme to show the key steps in the biosynthesis of penicillins and cephalosporins.

dihydrothiazine ring of the cephalosporin nucleus^{9,11}. The reaction requires dioxygen, and is catalysed by a 2-oxoglutarate-dependent ferrous enzyme. In prokaryotic cephalosporin producers, this enzyme is DAOCS¹², whereas in eukaryotic cephalosporin producers the reaction is catalysed by a bifunctional enzyme, deacetoxy/deacetylcephalosporin C synthase⁹. The prokaryotic and eukaryotic enzymes are closely related by sequence^{13,14}.

Here we report three high-resolution crystal structures for prokaryotic DAOCS from *Streptomyces clavuligerus*: the structure of the iron-free apoenzyme, the enzyme complexed with Fe(II), and the structure of the complex with Fe(II) and 2-oxoglutarate. The results imply a mechanism by which the enzyme-Fe(II) complex reacts with 2-oxoglutarate and dioxygen to give the reactive ferryl species, a process common to many non-haem oxygenases. Other

notable 2-oxoacid-dependent ferrous enzymes are prolyl hydroxylase, which is involved in collagen biosynthesis; gibberellin 3 β -hydroxylase, a mutation of which influences stem length in plants and was one of the traits studied by Mendel¹⁵; 1-aminocyclopropane-1-carboxylate oxidase (the ethylene-forming enzyme); and clavaminic acid synthase, which is involved in the biosynthesis of the β -lactamase inhibitor, clavulanic acid. Within the family of 2-oxoacid-dependent enzymes, DAOCS belongs to a subfamily^{1,10,14,15} the members of which show sequence similarity with IPNS, an enzyme that does not use a 2-oxoacid in catalysis.

The iron-free form of DAOCS crystallizes in space group R3 as a crystallographic trimer. The main chain of the protein folds into a conserved jelly roll with flanking helices. Three surface loops are disordered (Table 1). An intersubunit contact within the crystal-

Table 1 Data collection, phasing and refinement statistics

Data collection and phasing	Resolution (Å)	Number of reflections	Unique reflections	Completeness (outer shell)	Twinning fraction	R_{merge}^* (outer shell)	Number of sites	Phasing power†
Native 1 $\ddagger$ (apoenzyme)	1.3 (1.33–1.30)	240,321	71,652	98.0% (98.5%)	0.141	0.040 (0.184)	–	–
Native 2§ (apoenzyme)	1.9 (1.93–1.90)	52,947	20,176	85.5% (49.0%)	0.131	0.038 (0.139)	–	–
SeMet (λ = 0.689 Å)	2.7 (2.8–2.7)	40,081	8,039	98.5% (98.8%)	0.057	0.044 (0.063)	–	–
SeMet (λ = 0.979 Å)	2.7 (2.8–2.7)	33,192	7,996	97.6% (91.8%)	0.057	0.045 (0.059)	5	3.40††
K ₂ Pt(CN) ₄ §	2.6 (2.69–2.60)	46,378	9,279	100.0% (99.8%)	0.177	0.046 (0.099)	1	1.23#
Xe§	2.1 (2.18–2.10)	75,148	16,409	93.8% (63.1%)	0.271	0.052 (0.093)	1	2.81#
DAOCS-Fe(II) complex‡	1.5 (1.53–1.50)	193,504	45,516	95.4% (95.0%)	0.245	0.051 (0.296)	–	–
DAOCS-Fe(II)-2-oxoglutarate complex☆	1.5 (1.53–1.50)	180,561	48,012	97.6% (99.7%)	0.145	0.054 (0.256)	–	–
Structure	Resolution (Å)	Number of reflections used	Residues in model	Number of solvent molecules	r.m.s. bond lengths	r.m.s. angles/dihedrals	R_{cryst}	R_{free}
Native 1 $\ddagger$ (apoenzyme)	20–1.3	66,990	1–80, 98–166, 178–248, 253–311	260	0.022	2.3/25	0.129	0.150
DAOCS-Fe(II) complex‡	20–1.5	41,474	2–79, 98–164, 180–248, 258–308	182	0.010	2.2/26	0.124	0.149
DAOCS-Fe(II)-2-oxoglutarate complex☆	20–1.5	42,718	1–82, 91–165, 178–195, 201–247, 257–309	193	0.009	2.1/26	0.125	0.156

* $R_{\text{merge}} = \sum_i \sum_h |I_{hi} - \langle I_h \rangle| / \sum_i \sum_h I_{hi}$, where I_{hi} = intensity of an individual measurement, $\langle I_h \rangle$ = mean intensity of the reflection.

† The phasing power, a derivative is defined as the ratio of the amplitude of the r.m.s. (root mean square) heavy atom scattering factor to the r.m.s. lack of closure.

‡ Data collected on a 345-mm MAR image plate detector at BM14 of the ESRF (λ = 0.80 Å).

§ Data collected on a Rigaku R-axis image plate detector with CuK α radiation in Uppsala.

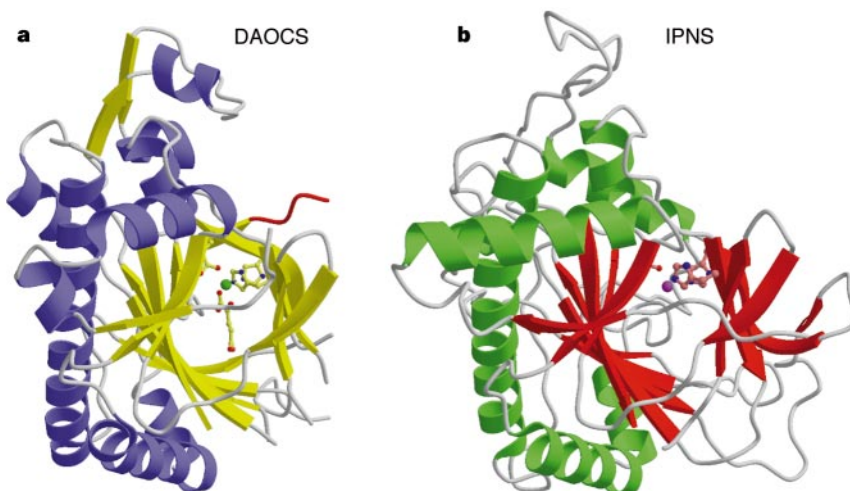
|| The two SeMet data sets were collected on the same crystal, using a CCD detector at BM14 of the ESRF.

†† Calculations based on dispersive differences against SeMet (λ = 0.689 Å).

Refined against Native2 and SeMet (λ = 0.689 Å).

☆ Data collected on a 300-mm MAR image plate detector at the MAX II synchrotron in Lund, Sweden (λ = 0.80 Å).

Figure 2 Comparison of the structures of deacetoxycephalosporin C synthase (DAOCS) and isopenicillin N synthase (IPNS). DAOCS (**a**) is a 2-oxoacid-dependent enzyme, whereas IPNS (**b**) is not. The two sequences show only 19% identity. Major departures from the IPNS structure can be found in the amino-terminal loop, in the region between residues 272 and 292 where two strands and a helix are inserted, in the active site and its environment, and at the C-terminal arm (highlighted in red in DAOCS).



lographic trimer is maintained by residues 308–311 (the carboxy-terminal arm, highlighted in Fig. 2a) in the apoenzyme. This arm penetrates the active site of the neighbouring subunit in a cyclic fashion, so that Lys 310 from one subunit lines the active-site pocket of the neighbouring subunit. Shortening the polypeptide chain from the C terminus by six residues diminishes activity. Together with the intricate network of interactions between symmetry-related subunits, this suggests that the trimeric structure of the apoenzyme found in the crystal might have some functional significance, for example in the protection of the C-terminal arm (rich in lysines and arginines) against proteases in the apo- or resting forms of DAOCS. However, DAOCS dissociates into monomers in the presence of Fe(II) and 2-oxoglutarate.

Figure 3 shows the geometry of the active site of DAOCS. Soaking crystals of the apoenzyme in a solution of Fe(II) results in iron binding (Fig. 3b). This causes only small structural changes; however, in agreement with the observation that the monomer is the catalytically competent form, the protruding C-terminal arm of the neighbouring molecule moves away from the active-site area, loosening intermolecular contacts. Careless addition of Fe(II) breaks up the crystals. The ferrous iron is ligated by His 183, His 243, Asp 185 and three solvent molecules in an almost perfect octahedral geometry (Fig. 3b). The coordination sphere of iron in other mononuclear ferrous enzymes similarly contains two histi-

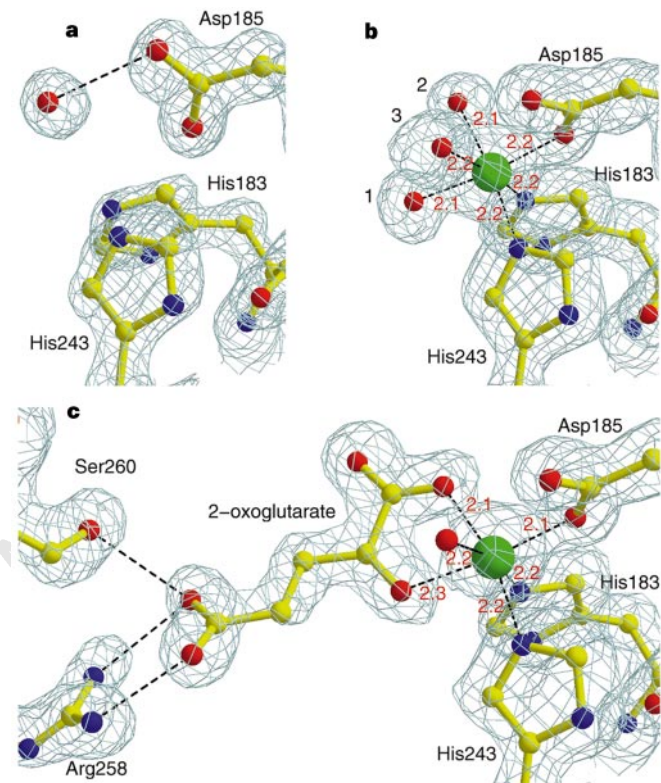


Figure 3 Structure of the active site in DAOCS. The site is shown in **a**, the apoenzyme; **b**, in the enzyme reconstituted with Fe(II); and **c**, in the Fe(II)-2-oxoglutarate complex. Structures for **b** and **c** were obtained under anaerobic conditions. The ferrous iron is ligated by Asp 185, His 183, His 243 and three solvent molecules (labelled 1, 2 and 3) in **b**. Solvent molecules 1 and 2 are replaced by 2-oxoglutarate in **c**. The co-substrate binds in a bidentate manner, with the 2-oxo-group *trans* to Asp 185. This site corresponds to the putative dioxygen binding site in IPNS. The 5-carboxylate of 2-oxoglutarate binds to a conserved RXS motif. A similar RXS motif binds the valine carboxylate of ACV in IPNS². The most likely binding site for dioxygen is opposite His 183, occupied by a solvent molecule in **c**. Electron density around this solvent molecule is similar to that in **b**, but, to avoid obstruction, was not drawn here (**c**). Iron-ligand distances (Å) are given in red.

dines and a carboxylate^{3,10,16}. In the ternary DAOCS-Fe(II)-2-oxoglutarate complex (Fig. 3c), the co-substrate binds in a bidentate manner, replacing two solvent molecules around the iron. The 1-carboxylate of 2-oxoglutarate binds *trans* to His 243, and the 2-carbonyl group binds *trans* to Asp 185. The latter site is equivalent to the position where nitric oxide, a dioxygen analogue, binds in IPNS (ref. 2). The stereochemistry of the co-substrate binding is similar to that proposed from chemical and mutagenesis studies for prolyl 4-hydroxylase¹⁷. The only accessible site around the iron in the 2-oxoglutarate complex of DAOCS is located in a hydrophobic pocket lined by residues Ile 192, Val 262, Phe 225, Phe 264 and the methyl group of Thr 190. This site is occupied by a solvent molecule in the structure (Fig. 3c), and we propose that this site is the binding site for dioxygen in DAOCS and related 2-oxoacid-dependent ferrous enzymes. The bound 2-oxoglutarate is flanked by the side chains of Met 180 and Val 262; the sulphur atom of Met 180 is in van der Waals contact with the C2 carbon atom of 2-oxoglutarate, whereas Val 262 maintains contact with the C3 carbon atom from the opposite side. The 5-carboxylate of the co-substrate forms a salt bridge with Arg 258 and is hydrogen bonded to Ser 260 (Fig. 3c). These residues are conserved within the DAOCS subfamily of enzymes.

Despite the conserved structural motifs in IPNS and DAOCS (Fig. 2), there is a significant difference in their use of dioxygen (Fig. 1). Whereas IPNS uses the full four-electron reduction of

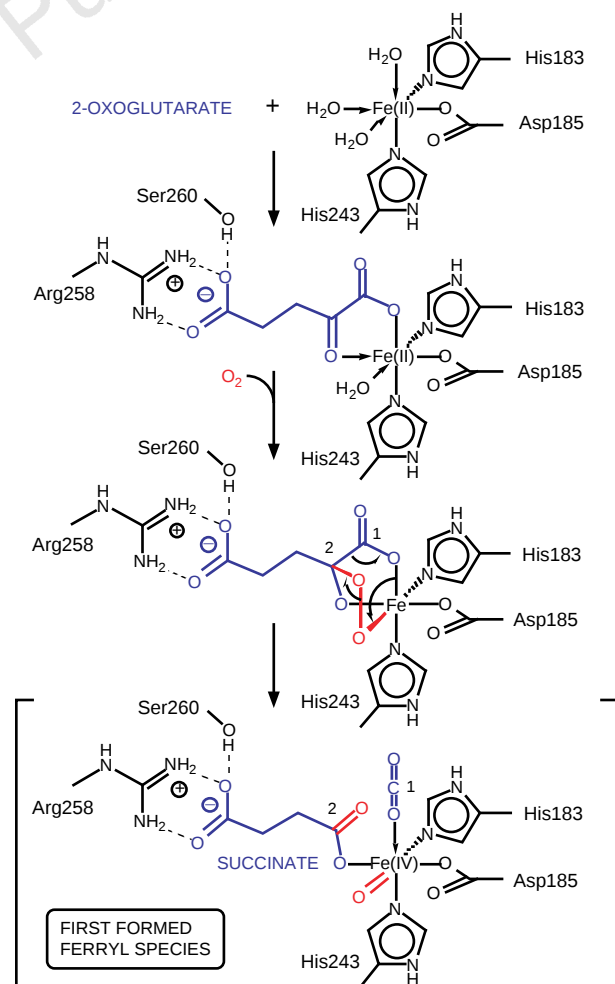


Figure 4 Scheme to show the proposed mechanism for the formation of the ferryl intermediate in DAOCS and related 2-oxoacid-dependent ferrous enzymes. Although the process is drawn as a concerted reaction in the last step, a stepwise process is also possible. The model does not exclude the presence of the penicillin substrate during ferryl formation.

molecular oxygen to water to create the strained ring structure of the penicillin core, DAOCS, like other 2-oxoglutarate-dependent ferrous enzymes, performs only a two-electron oxidation of its substrate (penicillin N in this case) for each molecule of dioxygen consumed. The ferryl form of IPNS is created by splitting dioxygen in a reaction coupled to the two-electron oxidation of the substrate, ACV, which probably involves closure of the β -lactam ring². However, the ferryl form of DAOCS is created by splitting dioxygen in a reaction coupled to the oxidative decarboxylation of the co-substrate, 2-oxoglutarate (Fig. 4). After these initial steps, both enzymes use their ferryl state to complete the remaining parts of their reactions. An important role of the protein in these reactions is to provide a shelter to reactive intermediates, and a possible function of the C-terminal arm of DAOCS and IPNS is to isolate reactive intermediates during catalysis.

The high reactivity of iron–dioxygen and iron–oxo intermediates can give rise to unwanted side reactions and may lead to the destruction of such enzymes¹⁸. The control of the reactivity of intermediates is therefore an important aspect of catalysis by these enzymes. Our structural results give indications of how control may be exercised in 2-oxoacid-dependent enzymes. In the ferrous form of the enzyme (Figs 3b and 4), the iron is ligated by three protein ligands and three solvent molecules. Binding of the co-substrate introduces a negative charge into the coordination sphere of the iron, thereby activating it for productive dioxygen binding³. The presence of the co-substrate thus gives the signal for the start of the reaction. Oxidative decarboxylation of 2-oxoglutarate maintains the negative charge around the iron, but the position of the carboxylate ligand shifts from opposite His 243 to opposite Asp 185 (Fig. 4, last step) as 2-oxoglutarate is converted to succinate. An electron-donating carboxylate ligand has a stabilizing effect on the ferryl species, whereas the shift in the position of the carboxylate during the reaction may have a role in directing the ferryl in later steps. The other reaction product, carbon dioxide, could either leave the iron or react with water *in situ* to give bicarbonate (or carbamate, through a reaction with an amine). This reaction could introduce a further negative charge into the environment of the ferryl. We note that some 2-oxoacid-dependent ferrous enzymes are activated by carbon dioxide¹⁸. The scheme outlined above could ensure that the reactivity of the ferrous iron is first enhanced towards dioxygen by the negative charge of the co-substrate, while the reactivity of the ferryl intermediate is tuned down by the charge of the products. The ferryl species may thus be stored safely within the protein to await a reaction with the 'main' substrate (penicillin N in this case). Controlled changes in the coordination sphere of the ferryl after substrate binding could modify its reactivity and ensure selective oxidation. A detailed understanding of how these processes occur may assist in the design of non-protein-based oxidative catalysts using a ferryl intermediate.

Many members of the family of mononuclear ferrous enzymes catalyse reactions that are synthetically impossible at present. Our results, together with sequence data^{1,14}, show that members of the DAOCS subfamily have similar three-dimensional structures (Fig. 2). The diversity in catalytic specificities of these enzymes on very similar structural platforms suggests that altering substrate and product specificities should be possible in the laboratory. This could be used for the synthesis of new antibiotics to combat the growing threat of resistant microorganisms. □

Methods

Crystallization and X-ray data collection. The DNA fragment that contained the DAOCS gene from *S. clavuligerus*^{13,19} was subcloned into the pET 11a vector and overproduced in *Escherichia coli* (M.D.L. *et al.*, in preparation). The purified protein was crystallized from 1.75 M ammonium sulphate, 5 mM 2-oxoglutarate and 0.1 M HEPES, pH 7.0–7.5. Crystals belong to space group R3, with unit-cell dimensions of $a = b = 106.4 \text{ \AA}$, $c = 71.2 \text{ \AA}$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$ (in the hexagonal setting), and one monomer in the asymmetric unit. They

diffract X-rays beyond a resolution of $1.3 \text{ \AA}$. All crystals were merohedral twins, with twinning ratios between 0.06 and 0.45. Data were collected at 90 K on station BM14, ESRF, Grenoble, and on beamline I-711, MAX-Lab, Lund, Sweden (Table 1).

Structure determination. The structure of the enzyme in its apoform was determined by multiple isomorphous replacement (Table 1) and by a space group general detwinning strategy that was based on an extension and modification of ref. 20. Details of this procedure will be published elsewhere (A.C.T.v.S. *et al.*, in preparation). Data were processed and scaled with HKL²¹ and CCP4 suites²². After detwinning, an initial phase set was calculated that could be extended to a resolution of $1.3 \text{ \AA}$ (Table 1). The procedure included solvent flipping by SOLOMON²³, and the subsequent construction of a free-atom model of DAOCS as described for the wARP procedure²⁴. This model was refined by ARP²⁵ and REFMAC²⁶. At the end of this procedure, R_{cryst} (defined as $|\Sigma|F_{\text{obs}}| - |\Sigma|F_{\text{calc}}|| / |\Sigma|F_{\text{obs}}| \times 100$) dropped from 50.7% for the initial free-atom model to 18.5% for the final free-atom model. Subsequent automatic chain tracing (A.P. and V.S. Lamzin, unpublished data) allowed 262 residues out of the 311 residues of the DAOCS molecule to be built in a single step. Electron-density maps were inspected and further interpreted with the program O²⁷, and the new models were refined with XPLOR²⁸ and SHELX97 (ref. 29). Quality checks with programmes from the Uppsala Software Factory (<http://alpha2.bmc.uu.se/usf/>) showed excellent geometry and a good agreement between models and data.

Structure determination of DAOCS complexes with Fe(II) and Fe(II)-2-oxoglutarate. Crystals of the apoenzyme were soaked at 15°C for 30–60 min in solutions containing 50 mM Fe(II) or 20 mM Fe(II) and 100 mM 2-oxoglutarate in an artificial mother liquor solution. Anaerobicity (0.1 p.p.m. oxygen) was maintained throughout, and crystals were flash frozen inside the anaerobic box. Initial models were obtained by rigid-body refinement using the structure of the apoenzyme (except for residues 307–311), and were then further refined with REFMAC²⁶ and SHELX97 (ref. 29) (Table 1).

Figures. Figures were drawn with a modified version of MOLSCRIPT³⁰.

Received 27 March; accepted 19 June 1998.

- Roach, P. L. *et al.* The crystal structure of isopenicillin N synthase, first of a new structural family of enzymes. *Nature* **375**, 700–704 (1995).
- Roach, P. L. *et al.* Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827–830 (1997).
- Hegg, E. L. & Que, L. The 2-His-1-carboxylate facial triad—an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur. J. Biochem.* **250**, 625–629 (1997).
- Brotzu, G. Recherche su di un nuovo antibiotico. *Lavori Istituto Igiene Cagliari* **1**–11 (1948).
- Newton, G. G. F. & Abraham, E. P. Cephalosporin C, a new antibiotic containing sulphur and D- α -aminoadipic acid. *Nature* **175**, 548 (1955).
- Abraham, E. P. & Newton, G. G. F. The structure of cephalosporin C. *Biochem. J.* **79**, 377–393 (1961).
- Hodgkin, D. C. & Maslen, E. N. The X-ray analysis of the structure of cephalosporin C. *Biochem. J.* **79**, 393–402 (1961).
- Abraham, E. P. & Newton, G. G. F. A comparison of the action of penicillinase on benzylpenicillin and cephalosporin N and the competitive inhibition of penicillinase by cephalosporin C. *Biochem. J.* **63**, 628–634 (1956).
- Baldwin, J. E. & Schofield, C. J. in *The Chemistry of β -lactams* (ed. Page, M. I.) 1–78 (Blackie, London, 1992).
- Schofield, C. J. *et al.* Proteins of the penicillin biosynthesis pathway. *Curr. Opin. Struct. Biol.* **7**, 857–864 (1997).
- Yoshida, M. *et al.* Cell-free expansion of penicillin N to deacetoxycephalosporin C by *Cephalosporium acremonium* CW-19 and its mutants. *Proc. Natl Acad. Sci. USA* **75**, 6253–6257 (1978).
- Dotzlauf, J. E. & Yeh, W. K. Purification and properties of deacetoxycephalosporin C synthase from recombinant *Escherichia coli* and its comparison with the native enzyme purified from *Streptomyces clavuligerus*. *J. Biol. Chem.* **264**, 10219–10227 (1989).
- Kovacevic, S., Weigel, B. J., Tobin, M. B., Ingolia, T. D. & Miller, J. R. Cloning, characterization, and expression in *Escherichia coli* of the *Streptomyces clavuligerus* gene encoding deacetoxycephalosporin C synthetase. *J. Bacteriol.* **171**, 754–760 (1989).
- Prescott, A. G. A dilemma of dioxygenases (or where biochemistry and molecular biology fail to meet). *J. Exp. Bot.* **44**, 849–861 (1993).
- Lester, D. R., Ross, J. J., Davies, P. J. & Reid, J. B. Mendel's stem length gene (*Le*) encodes a gibberellin 3- β -hydroxylase. *Plant Cell* **9**, 1435–1443 (1997).
- Landman, O., Borovok, I., Aharonowitz, Y. & Cohen, G. The glutamine ligand in the ferrous iron active site of isopenicillin N synthase of *Streptomyces jumoningensis* is not essential for catalysis. *FEBS Lett.* **405**, 172–174 (1997).
- Myllyharju, J. & Kivirikko, K. I. Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *EMBO J.* **16**, 1173–1180 (1997).
- Zhang, Z.-H., Barlow, J. N., Baldwin, J. E. & Schofield, C. J. Metal-catalyzed oxidation and mutagenesis studies on the iron(II) binding site of 1-aminocyclopropane-1-carboxylate oxidase. *Biochemistry* **36**, 15999–16007 (1997).
- Morgan, N. *et al.* Substrate specificity of recombinant *Streptomyces clavuligerus* deacetoxycephalosporin C synthase. *Bioorg. Med. Chem. Letts.* **4**, 1595–1600 (1994).
- Fisher, R. G. & Sweet, R. M. Treatment of diffraction data from protein crystals twinned by merohedry. *Acta Crystallogr. A* **36**, 755–760 (1980).
- Otwinowski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. W. & Bailey, S.) DL/SCI/R34, 55–62 (Daresbury Laboratory, Warrington, UK, 1993).
- The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

23. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATP-ase. *Acta Crystallogr. D* **52**, 30–42 (1996).
24. Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D* **53**, 448–455 (1997).
25. Lamzin, V. S. & Wilson, K. S. Automated refinement of protein molecules. *Acta Crystallogr. D* **49**, 129–147 (1993).
26. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
27. Jones, T. A., Bergdoll, M. & Kjeldgaard, M. in *Crystallographic and Modelling Methods in Molecular Design* (eds Bugg, C. & Ealick, S.) 189–190 (Springer, New York, 1990).
28. Brünger, A. T., Kuriyan, J. & Karplus, M. Crystallographic R-factor refinement by molecular dynamics. *Science* **235**, 458–460 (1987).
29. Sheldrick, G. M. & Schneider, T. R. SHELXL: High-resolution refinement. *Methods Enzymol.* **277**, 319–343 (1997).

30. Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132–134 (1997).

Acknowledgements. We thank Å. Danielsson and R. Bhikhabhai (Pharmacia Biotech, Uppsala) for improved protein purification procedures; G. Larsson and A. M. Vålgård for bigger and better crystals; D. van der Spoel, C. M. R. Wouts, E. Wikman and G. Kleywegt for discussion; C. Andersson for in-house X-ray facilities; V. Biou, Z.-H. Zhang, P. L. Roach, J. Keeping and J. Pitt for their help, and Y. Cerclius and A. Svensson for data collection facilities, at MAX-Lab in Lund; and A. Dahl for a stimulating environment. This work was supported by EU-BIOTECH and the Swedish Research Councils, MFR and NFR. The Oxford Centre for Molecular Sciences is supported by BBSRC and MRC.

Correspondence and requests for materials should be addressed to I.A. (e-mail: inger@xray.bmc.uu.se) or C.J.S. (e-mail: christopher.schofield@chemistry.oxford.ac.uk). Coordinates and structure factors have been deposited with the Protein Data Bank (entries 1dcs and r1dcsf for the apoenzyme, 1rxg and r1rxsf for the Fe(II) complex, and 1rxg and r1rxsf for the Fe(II)-2-oxoglutarate complex). These entries will be held for one year.

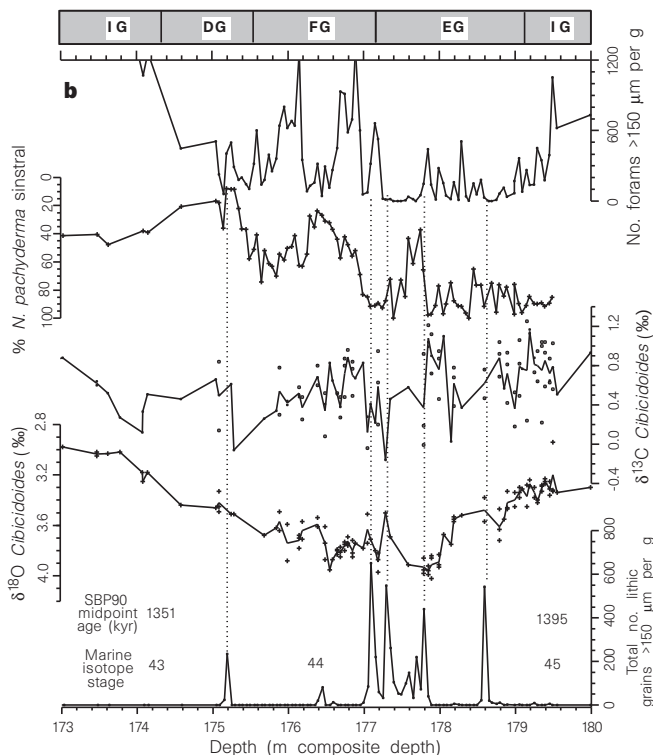
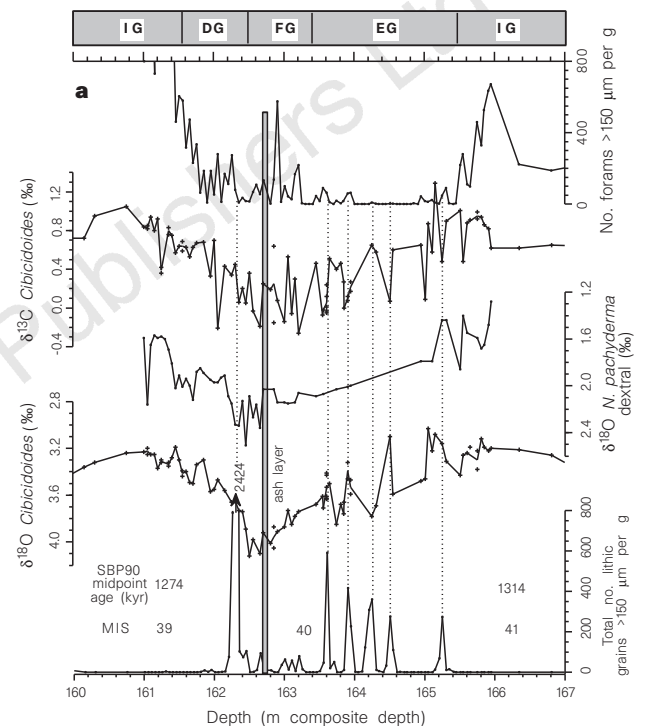
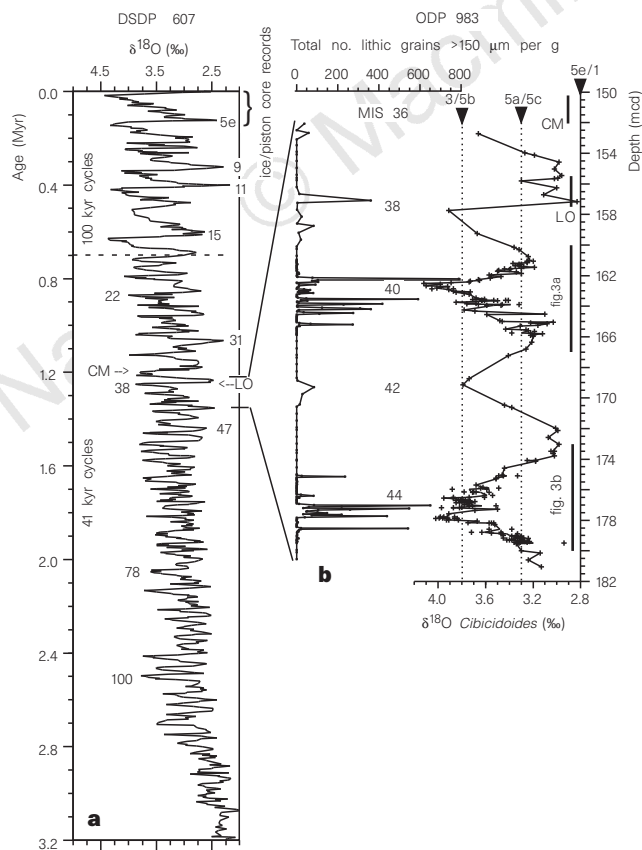
erratum

Millennial-scale climate instability during the early Pleistocene epoch

M. E. Raymo, K. Ganley, S. Carter, D. W. Oppo & J. McManus

Nature **392**, 699–702 (1998)

Figures 2 (below) and 3 (right) of this Letter contained some typographical errors. The corrected figures are reproduced here. □



Peptide and protein paraphernalia

The equipment and reagents in this week's selection of new products are biased towards protein preparation and handling. **Melanie II**, **profiBlot II** and **NMR X-Plor** feature.

ReadyPreps protein preparation kit

From Epicentre Technologies

Formulated to be a non-denaturing method of protein isolation, especially recombinant gene products expressed in Escherichia coli

The ReadyPreps kit can prepare small samples rapidly — taking as little as 20 minutes — for in-process monitoring of expression, for screening large numbers of transformants for activity, or for larger-scale production of proteins. By assaying the resulting lysates for activity, the kit allows for quick determination of whether functional products are being produced. The ReadyPreps kit is compatible with most downstream purification methods, including DNA precipitation, ammonium sulphate precipitation, anionexchange and sulphate precipitation, anion exchange and affinity chromatography, and tangential-flow filtration.

Reader Enquiry No. 100

ProtoPrep

From National Diagnostics

Meltable synthetic electrophoresis matrix for the one-step purification of proteins

ProtoPrep is a meltable, synthetic gel matrix intended as an acrylamide replacement for the separation, purification and recovery of proteins. ProtoPrep can be used to separate proteins from 8 to 220K. It can be used to obtain usable samples by simply excising and melting the relevant gel section. The gel will melt only once and thereafter remain liquefied even at room temperature. The protein of interest may then be processed directly in the aqueous phase without having to be purified away from the inert gel matrix.

Reader Enquiry No. 101

PAGEr

From FMC

Precast polyacrylamide minigels for protein electrophoresis

These gels are designed to provide a fast, convenient and reproducible means of separating native and denatured proteins over a wide molecular weight range. The gels fit most vertical minigel boxes and are available in 8 × 10 and 10 × 10 cm formats. According to the manufacturer, the system includes several innovative enhancements, including a novel cassette design that opens easily using the enclosed comb. PAGEr gels are presently available with the Tris-glycine



PAGEr: separations in standard minigel boxes.

buffer system and can be ordered in concentrations of 7.5, 10, 12, 15 per cent, or gradient concentration ranges of 40–20 and 10–20 per cent.

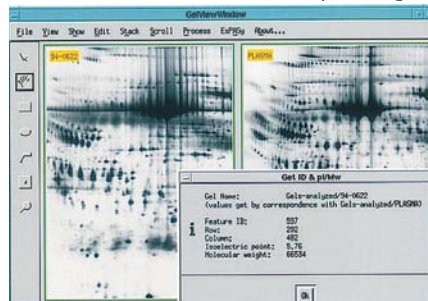
Reader Enquiry No. 102

Melanie II

From Bio-Rad

Two-dimensional PAGE software that helps speed pI and MW evaluations

This gel-analysis package enables faster cross-referencing with individual or worldwide standard references. Arduous tasks, such as gel matching, stacking, cross-referencing, isoelectric point (pI) and molecular weight (MW) determination, can now be carried out more quickly and with accuracy using this package that has been designed for spot identification, quantification and comparisons. By clicking on toolbar icons, the software can detect, group and quantify specific features and align gels. Drop-down menus and macros can be tailored to individual applications for functions such as creating synthetic reference gels or matching to stored results. Melanie II also enables users to seamlessly access an Internet database of master gels, which is constantly updated, for cross-referencing experimental results to a worldwide library of standard organisms and DNA preparations. Supported on Power Mac and Windows95 platforms, the software has options for single- and high-volume, two-dimensional analysis. Report



Look what Bio-Rad have done to PAGE software.

generation can be customized for each application, and specific information saved or printed during analyses.

Reader Enquiry No. 103

profiBlot II

From Tecan

Automated western blotting designed to confirm HIV and other viruses efficiently

This is built as a 'walk-away' system that ensures reproducible results and improves laboratory safety. Dispensing and wash volumes can be varied from 250 µl to 3,000 µl; incubation times can be set from 1 min to 999 min. There are five shaking speeds, and up to 36 strips can be processed on one tray. The system uses simple Windows-based PC software, or on-deck keyboard programming, and permits the storage of 20 protocols, each with up to 60 steps. This system has a semiautomatic calibration method that allows individual pumps to self-check and recalibrate against a specific liquid. profiBlot II comes in two versions: the N version has two reagent channels for fast dispensing and four channels for single-position reagent addition. Three aspirating needles are used to remove liquid from the tray; with the S version, six channels are used for normal reagent addition, and liquid is drained from the tray into a waste channel. The system should prove useful for routine clinical and research laboratories required to identify HIV, HTLV, HCV or Lyme disease pathogen.

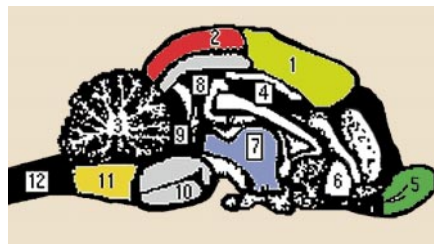
Reader Enquiry No. 104

Brain Blot

From Chemicon

Study protein distribution in the brains of rats and mice

Chemicon has carefully dissected and processed 12 anatomically and functionally distinct areas of adult rat and adult mouse brain for the study of proteins using western blotting. These areas include the frontal cortex (labelled 1 in graphic over page), posterior cortex (2), cerebellum (3), hippocampus (4), olfactory bulb (5), striatum (6), thalamus (7), midbrain (8), entorhinal cortex (9), pons (10), medulla (11) and spinal cord (12). These brain proteins have been extracted in standard Laemmli SDS-PAGE buffer (reduced) and electrophoresed on 4–14% gradient gels. The proteins are then electro-blotted onto nitrocellulose and blocked with PBS/milk. A lane of pre-stained molecular weight markers is included in each blot to assist users in identifying the size of



Proteins on your mind? Chemicon can find them.

the proteins. Brain Blot, or any other blot, may be reused many times by stripping antibodies at room temperature for 5–15 min using the Re-Blot western blot recycling (no heating or mercaptoethanol is required).

Reader Enquiry No. 105

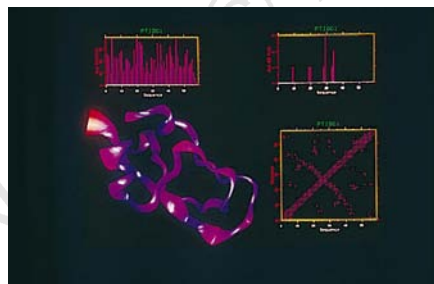
NMR X-Plor

From Molecular Simulations

This program extends Insight II's NMR structure refinement capabilities

NMR X-Plor links the graphical facilities of Insight II with the structure determination protocols of X-Plor/DG, a program used for macromolecular structure determination. This software is commonly used for integrated NMR and X-ray crystallography, allowing users to read and edit NMR-derived restraints, set up calculations and analyse the resulting structures. NMR X-Plor's menu driven interface simplifies the use of X-Plor/DG's structure determination protocols, while also allowing customization by more advanced X-Plor users.

Reader Enquiry No. 106



X-Plor MSI combines NMR and crystallography.

Labelling, detection and identification

GFP/YFP filter

From Glen Spectra

New GFP/YFP filter sets for flowcytometry

A filter set for the latest flow cytometry technique has been developed by Omega Optical. Glen Spectra offers the GFP/YFP flow filter set for sorting and analysing cells expressing green fluorescent protein, yellow fluorescent protein, and similar mutants. Using the 488-nm laser line, both GFP and YFP can be excited simultaneously and the filter set will allow both emissions to be separated by a dichroic, which passes the GFP signal to one PMT and reflects the YFP signal to a second PMT. The respective signals are then cleaned up using two emission filters in front of the PMTs.

Reader Enquiry No. 107

ProteinLynx literature

From Micromass

New literature introduces this automated protein identification system for high-throughput proteomics

This system has been designed to combine sample management and automated mass spectrometry data acquisition with client-server bioinformatics to create a single, integrated solution. ProteinLynx is implemented on a secure Windows NT platform, an extension of 'MassLynx' technology. The standard analyser options include MALDI-TOF MS workstations for rapid primary mass mapping and electrospray MS-MS workstations for sequence tag determination or *de novo* sequencing. ProteinLynx is compatible with FASTA-formatted databases, enabling automated hierarchical search strategies to be implemented on both local and remote servers. Furthermore, the system facilitates linking to host bioinformatics systems. ProteinLynx uses a three-tiered analytical strategy for protein identification. The primary screen employs a MALDI-TOF MS workstation to conduct automatic peptide-mass fingerprinting of very large arrays of

digested protein samples that have been excised from two-dimensional gels. The secondary screen involves sequence tagging — a dissolved protein digest sample is analysed using a Q-TOF LC-MS-MS workstation, resulting in full product ion spectra being recorded. Subsequent searching may be achieved with ProteinProbe as ProteinLynx incorporates a 'find sequence tag' tool to generate the 'tag' automatically. Proteins and peptides that remain unidentified by this technique are fully sequenced by ES MS-MS. The third stage of analysis, *de novo* sequencing, also involves Q-TOF's quadrupole-TOF technology, which is said to provide sub-femtomole quantities of peptides. PepSeq, a new multi-layered peptide sequencing tool, has been designed to yield full product ion (MS-MS) spectra, which are initially processed by charge state de-encryption, identification of C-terminally tagged and sequence tags established.

Reader Enquiry No. 108

Luciferase-luminometer tandem

From Dynex Technologies and Promega

Combining Promega's Dual-Luciferase (DLR) reporter assay with Dynex's MLX microtitre plate luminometer technology

Dual-Flash software rounds out this package and has been designed specifically for the DLR assay. The DLR assay utilizes both the firefly (*Photinus pyralis*) and the *Renilla reniformis* (sea pansy) luciferases. Both enzyme assays are performed in a single tube, or well, and produce a linear 'glow-type' signal with sub-attomole sensitivities. The firefly luciferase is used to measure the 'experimental' sample while the *Renilla* luciferase is used as a control against which the experimental sample can be normalized. The DLR 1000 assay system is configured for use in 96-well plate reading luminometers equipped with two injectors (a 100-assay system is also available). The MLX microtitre plate luminometer is designed for clinical research and high-throughput screening/drug discovery. The electronic

ADVERTISEMENTS

PEPTIDE MAKER

Custom peptides, conjugates and antibodies

Unrivalled experience and expertise with assured quality

For more information on AMRI's Peptide and Protein Research Services, please contact:

AMRI Research Products Ltd., Marshhead Castle, Exeter, EX8 8HD, UK
Tel: +440 1526 681010; Fax: +440 1526 681080; URL: <http://www.amri-research.com>

READER ENQUIRY NO. 32

ANTIBODIES

For Research & Routine Histology

Information - Angiogenesis - Tumors
Leukocytes - Endothelial Cells
Pathological Markers

Custom Monoclonals and Polyclonals

BMA Biomedicals Ltd.
(800) 634-3183 (US)
Int: 441 61 811 62 22
+41 61 811 60 06 (fax)
info@bma.ch

READER ENQUIRY NO. 33

design provides eight decades of dynamic range for measuring both bright and sensitive light output. Advanced electro-optics have been engineered to measure light with a reproducibility of less than one per cent CV and linearity greater than 99 per cent to provide. Each photomultiplier tube provides sub-attomolar luciferase sensitivity and instrument-to-instrument consistency that is stated to be within 10 per cent. The plate-carrier mask reduces well-to-well crosstalk down to 2.5×10^{-5} . The MLX has an external RS-232 serial communication port and an optional robotic plate carrier for compatibility with robotic systems. The instrument is available with up to three dispensers, two of which are required for the DLR assay. The MLX is driven by an on-board 75 MHz Intel Pentium microprocessor and controlled by Windows-base operating and data reducing software.

Reader Enquiry No. 109

AutoFluorescent Proteins

From Quantum Biotechnologies
Fluorescent tags and reporter systems

Quantum's autofluorescent proteins (AFPs) are variants of the fluorescent light emission protein in the jellyfish *Aequorea victoria*. Modifications of the chromophore sequences of wtGFP have produced bright and efficient blue- or red-shifted fluorescent proteins. A variety of optimized constructs are now available for various applications. CMV-promoter-driven AFPs are well suited for transient expression in mammalian cells, as well as for the generation of AFP-fusion constructs or promoter evaluation. AFP-neo fusions, driven by either the murine PGK or Pol 11 promoters, offer dual-selection potential for the rapid evaluation of transfection efficiencies and the establishment of stably transformed fluorescent cell lines. According to the manufacturer, pQB163 and pQB167 use the powerful T7 promoter to generate very high expression levels of rsGFP and BFP, respectively, in bacteria. Quantum's BLUF fluorescent protein is recommended for use with their brilliant red-shifted green protein. Use of both proteins makes possible double-labelling experiments and allows real-time, multi-parameter analyses of gene expression by microscopy, FACS or fluorometry.

Reader Enquiry No. 110

Epitope tagging reagents

From Serotec

Monoclonal antibodies to the Pk His (polyhistidine), GST (Schistosoma japonicum glutathione S-transferase) and the MBP (maltose binding protein) epitope tags

These tagging reagents have been created for the identification of recombinant proteins. A major difficulty associated with the investigation of recombinant proteins is their identification in the absence of a suitable protein-specific monoclonal antibody. These antibodies are stated to be specific, simple to use, create reproducible results and are suitable for use in numerous applications. Such applications include western blot analysis, immunoprecipitation, immunofluorescence, ELISA, immunocytochemistry flow cytometry. These antibodies can also be used for purification, examination of intracellular transport and distribution of proteins and to detect protein interactions.

Reader Enquiry No. 111

Peptides

Bioactive peptides

From Sigma

The Sigma range includes over 40 new bioactive peptides for use in signal transduction and basic peptide research

Sigma bioactive peptides included such molecules as CTOP, SPARC fragment 119–122, orphanin FQ, urocortin, astressin, influenza hemagglutinin (HATag) peptide and c-erb-3 fragment 1265–1278. Also newly introduced are several workhorse peptides, including amyloid β -protein fragment 1–43, leuprolide and leptin fragment 22–56. With these new additions, the company now offers over 600 bioactive peptides in 20 categories. According to Sigma, these new peptides are competitively priced, meet the company's stringent quality control and purity requirements and are backed with comprehensive technical support. For added convenience, Sigma bioactive peptides are available in a variety of



Choose from 600+ bioactive peptides.

package sizes, including quantities suitable for laboratory and manufacturing use.

Reader Enquiry No. 112

Peptide synthesis service

From Genosphere Biotechnologies

Peptide synthesis using Fmoc solid-phase chemistry

For research, this service will produce a standard synthesis quantity of 15–40 mg of crude product. Using reversed-phase HPLC, four levels of purity are available, with all peptides confirmed for specifications and delivered with an analytical HPLC profile and a mass spectrum. Additional analyses are offered, including amino acid analysis. Moreover, an extensive range of modification, labelling and conjugation options are available: MAPs and protein conjugation, haptens, dyes and phosphorylation.

Reader Enquiry No. 113

Antibodies

Antibodies to yeast proteins

From Autogen Bioclear

A comprehensive range of antibodies to yeast proteins

These primary antibodies from Santa Cruz are raised against *Saccharomyces cerevisiae* and *S. pombe* proteins, such as those utilized in signalling, cell-cycle and DNA replication. Some 200 new antibodies are available to yeast cell-cycle proteins (CLN, CLB, CDC, SWEI, UBC9 and Nap1), cell-cycle checkpoint proteins (RAD, MAD, MEC POL, PDSI and DUN), cell-cycle transcription factors (SW14, SW16 and MBPI), histone acetylases and deacetylases (ADA, RPD, HDAI, SIN3 and GCN5) and telomere-associated proteins (TELI, RAPI, SIR and HDPI). Autogen Bioclear also offers antibodies to DNA replication factors, the pheromone response MAP kinase pathway; the osmoregulation MAP kinase pathway; the cell wall integrity MAP kinase pathway; and the RAS/cAMP pathway.

Reader Enquiry No. 114

ADVERTISEMENTS

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT

PROTEIN SEQUENCING
Membrane-bound, solid or solution samples
\$12 per cycle
5 cycles minimum - no set up charge

For more information please contact Dr. Jon Gray.
Phone/Fac: +44 (0)191 222 8612
Email: JON.GRAY@newcastle.ac.uk

DNA sequencing (ABI 377-5000, 377 Chromat)
also available at competitive prices

PEPTIDES
Synthesis & Specialist Chemistry
HF Cleavage
Insertion of amino acid analogues
Modified, labelled & Cyclic peptides
HPLC & electrospray mass analysis

Quotations: Peptide@DialPipe.com
Tel: 01457-875798 FAX: 01457-871088
Peptide Products Ltd (UK)

READER ENQUIRY NO. 42

READER ENQUIRY NO. 43

Antibodies to cytoskeletal proteins

From BABCO

Antibodies to help study how eukaryotic cells assume various shapes, provide communication pathways and direct movement on these protein filaments

This network, known as the cytoskeleton, is a highly dynamic structure directly responsible for muscle contraction, transportation of organelles, segregation of chromosomes during mitosis and cell shape changes during vertebrate embryo development, among other functions. BABCO provides antibodies specific for several cytokeratins (MK1, 5, 6, 10, 14, HK1 fillagrin and loricrin), non-muscle myosins (MHCA and MHCB) tubulins (β III and γ) kinesins (squid kinesin light chain, kinesin II and SUK-4) and kinesin-related proteins (HIPYR), and other motor proteins. This growing family of antibodies offers scientists useful tools for studying microtubule dynamics cell motility and the scaffolding of the cell as a whole.

Reader Enquiry No. 115

QuantiBrite

From Becton Dickinson

Quantitation-grade, custom 1:1, PE:monoclonal antibody conjugates

PE-conjugated monoclonal antibodies are for research use with the QuantiBrite fluorescence quantitation system. Fluorescence quantitation demands the use of high-quality PE conjugates and ensures accurate and reproducible results. Calibrating a flow cytometer with QuantiBrite PE beads gives estimates of the number of molecules of PE bound per cell. The more desirable unit, antibodies bound per cell (ABC), requires knowledge of the PE:mAb ratio of the conjugate. While a 1:1 ratio is often assumed



Quantitation-grade PE:mAb conjugates.

for ABC calculations, this is not necessarily the case. Many applications are emerging where quantitation is necessary. Combining 1:1 PE:mAb conjugates with QuantiBrite PE beads (and the soon-to-be-released QuantiCALC software) researchers should be able to perform this kind of quantitation reproducibly.

Reader Enquiry No. 116

Protein products

From Pierce

The company's offers a variety of protein research tools

Researchers looking for fusion protein products can now order a free copy of the company's fusion protein products brochure. This brochure provides information on coated/activated microwell plates, immobilized supports and detergent reagents to screen, detect, purify and cleave commonly used fusion proteins. These include glutathione S-transferase (GST), maltose binding protein (MBP), polyhistidine and biotin. Also available from the manufacturer is the SuperSignal western blotting substrates handbook with abbreviated and full-length western blot protocols, as well as recommendations for antibody concentration and blocking buffer optimization. An easy-to-follow troubleshooting guide explains possi-

ble causes and solutions, in the event that users encounter a problem with a substrate. Pierce's protein detection and quantitation brochure features new protein chemistry-based and immunochemistry-based products designed for proteomics, high-throughput screening (HTS) and immunoassay development. Featured items include total proteins, glycoprotein, antibody and fusion protein assay products, activated and pre-coated microwell plates, as well as blocking agents, detection reagents and substrates (chromogenic, fluorogenic and chemiluminescent). Also in the area of fusion protein purification is the company's B-PER bacterial protein extraction reagent. B-PER products offer a gentle method of complete bacterial cell lysis so there is no need to use sonication or any special instrumentation. Also, the B-PER reagent recovers both soluble and insoluble recombinant protein from *Escherichia coli*.

Reader Enquiry No. 117

The Antibody Source

From Biodesign

A catalogue featuring 4,000 monoclonal and polyclonal antibodies, purified antigens and assay development reagents

Products specific to adhesion molecules, apolipoproteins, apoptosis and assay development are available. This products should find use in the study of autoimmune diseases, cancer, cell receptors, collagens, conjugates, cytokines, growth factors, hemostasis, hormones, and many other areas.

Reader Enquiry No. 118

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

BIANDEM[®]
EQUILIBRIUM DIALYSIS SYSTEM

The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

BIANDEM
Dialyzer Box 500112
D. 87115 Mönster

READER ENQUIRY NO. 61

PIERCE
SuperSignal[®] LRA for
Luminescence-based Applications[®]
Delivers Femtogram-level Sensitivity for
Luminescence-based ELISAs

Features/Benefits:

- Ultrahigh femtogram-level sensitivity
- Low background and rapid processing
- Eight-hour working solution stability; etc. month room temperature reagent shelf life

Tel: 800-854-2725 or 813-854-2725 • Fax: 813-854-7318
info@piercelab.com • Internet: www.piercelab.com

• Patent Pending

READER ENQUIRY NO. 41

Check out
our homepage at
<http://www.resgen.com>

Peptides
\$18/residue

Antibodies
monoclonal - \$8,500
polyclonal - \$1,195

Complete services for your
custom peptide and polyclonal or
monoclonal antibody needs.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 37